

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 18, 2002, 11:08:05 ; Search time 83.16 Seconds
(without alignments)
5685.367 Million cell updates/sec

Title: US-09-800-198-8
Perfect score: 14581
Sequence: 1 MDVKDRRHSRLTRGRCREC.....ELADSSNIQFLRQNMGRK 2733

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues
Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_19.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phase.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_rvirus.*
16: sp_bacteriaph.*
17: sp_archaea.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	14316.5	98.2	2764	11 Q9WTS5	Q9WTS5 mus musculus
2	14291	98.0	2765	11 Q9RIK2	Q9RIK2 rattus norv
3	13934.5	95.6	2802	13 Q9DER5	Q9DER5 gallus gall
4	10309	70.7	2715	11 Q9WTS6	Q9WTS6 mus musculus
5	9522	65.3	2825	11 Q70465	Q70465 mus musculus
6	9518	65.3	2771	11 Q9WTS7	Q9WTS7 mus musculus
7	9424	64.6	2590	13 Q9W7R4	Q9W7R4 brachydanio
8	9241.5	63.4	2824	13 Q9W7R3	Q9W7R3 brachydanio
9	9209	63.2	2346	11 Q9JLC1	Q9JLC1 mus musculus
10	9055	62.1	1737	4 Q9ULC2	Q9ULC2 homo sapien
11	8541.5	58.6	2705	13 Q9W6V6	Q9W6V6 gallus gall
12	8484.5	58.2	2731	11 Q9WTS4	Q9WTS4 mus musculus
13	8406.5	57.7	2725	4 Q9UKZ4	Q9UKZ4 homo sapien
14	5822	46.8	1769	4 Q9P273	Q9P273 homo sapien
15	4848	33.2	930	11 Q9JLC0	Q9JLC0 mus musculus
16	4465	30.6	849	4 Q9N768	Q9N768 homo sapien

17	4393.5	30.1	1364	4	075981	075981 homo sapien
18	4060.5	27.8	831	13	Q9PU49	Q9PU49 gallus gall
19	3939.5	27.0	2731	5	061307	061307 drosophila
20	3933.5	27.0	2711	5	018366	018366 drosophila
21	3916.5	26.9	2515	5	024551	024551 drosophila
22	3915.5	26.9	2515	5	Q9VNU6	Q9VNU6 drosophila
23	3880.5	26.6	1045	4	Q9NVW1	Q9NVW1 homo sapien
24	3628.5	24.9	964	4	Q9NV77	Q9NV77 homo sapien
25	2661.5	18.3	730	4	Q96MS6	Q96MS6 homo sapien
26	2335	16.0	625	4	Q96SV2	Q96SV2 homo sapien
27	2212	15.2	2560	5	Q21980	Q21980 caenorhabdi
28	2090	14.3	609	11	Q923P3	Q923P3 mus musculu
29	2016	13.8	337	11	Q9RIK0	Q9RIK0 rattus norv
30	1582.5	10.9	272	11	Q9RIJ9	Q9RIJ9 rattus norv
31	1537.5	10.5	266	11	Q9RIK1	Q9RIK1 rattus norv
32	1384	9.5	229	11	Q9QVZ1	Q9QVZ1 mus musculu
33	1328	9.1	442	4	Q9NZJ2	Q9NZJ2 homo sapien
34	1203.5	8.3	560	4	Q98202	Q98202 homo sapien
35	1174	8.1	777	5	Q24550	Q24550 drosophila
36	1167	8.0	777	5	Q9VYP1	Q9VYP1 drosophila
37	1132	7.8	375	4	Q9P2P4	Q9P2P4 homo sapien
38	1110	7.6	587	5	Q9VYN8	Q9VYN8 drosophila
39	1043.5	7.2	1124	5	Q9VYN6	Q9VYN6 drosophila
40	752.5	5.2	278	13	Q9DEQ8	Q9DEQ8 gallus gall
41	726	5.0	1810	13	Q90824	Q90824 gallus gall
42	711.5	4.9	1714	13	Q90995	Q90995 gallus gall
43	675.5	4.6	2019	11	Q64706	Q64706 mus musculu
44	668.5	4.6	1532	13	Q90994	Q90994 gallus gall
45	632.5	4.3	4006	11	035452	035452 mus musculu

ALIGNMENTS

RESULT 1						
Q9WTS5	Q9WTS5	PRELIMINARY;	PRT;	2764	AA.	
ID	Q9WTS5	PRELIMINARY;	PRT;	2764	AA.	
AC	Q9WTS5;					
DT	01-NOV-1999 (TrEMBLrel. 12, Created)					
DT	01-NOV-1999 (TrEMBLrel. 12, Last sequence update)					
DT	01-DEC-2001 (TrEMBLrel. 19, Last annotation update)					
DE	TEN-M2.					
GN	ODZ2 OR TEN-M2.					
OS	Mus musculus (Mouse).					
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;					
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.					
OX	NCBI_TaxID=10090;					
RP	[1]					
RN	SEQUENCE FROM N.A.					
RC	STRAIN-BALB/C; TISSUE-BRAIN;					
RA	Oohashi T., Zhou X., Feng K., Richter B., Moergelin M., Perez M.T.,					
RA	Su W., Chiquet-Ehrismann R., Rauch U., Faessler R.;					
RT	"Mouse Ten-m/Odz is a new family of dimeric type II transmembrane					
RT	proteins expressed in many tissues.";					
RL	J. Cell Biol. 0:0-0(1999).					
DR	EMBL; AB025411; BAA77397.1; -					
DR	HSSP; P35555; IEMN.					
DR	MGD; MGI:1345184; Odz2.					
DR	InterPro; IPR000561; EGF-like.					
DR	InterPro; IPR002049; Laminin_EGF.					
DR	InterPro; IPR001258; NHL.					
DR	Pfam; PF01436; NHL; 2.					
DR	Pfam; PF01436; NHL; 2.					
DR	PRINTS; PR00011; EGF-LAMININ.					
DR	SMART; SM00181; EGF; 5.					
DR	PROSITE; PS00022; EGF_1; UNKNOWN_8.					
DR	PROSITE; PS01186; EGF_2; 7.					
KW	EGF-like domain; Glycoprotein.					
SO	SEQUENCE 2764 AA; 306464 MW; 73BA3D916D0F0344 CRC64;					

Query Match 98.2%; Score 14316.5; DB 11; Length 2764;
Best Local Similarity 96.8%; Pred. No. 0;

Matches 2684; Conservative 23; Mismatches 16; Indels 51; Gaps 3;

Qy 1 MDVKDRRHSRLTRGCKGKCRVTTSSLDSEDCRVPTQKYSSESSETLKAYDHDHSMHYGNR 60
Db 1 MDVKDRRHSRLTRGCKGKCRVTTSSLDSEDCRVPTQKYSSESSETLKAYDHDHSMHYGNR 60
Qy 61 VTDLTHRESDEFPROGNTFLAELGICEPSPHRSYCSDMGILHOGYSLSGSDADSDE 120
Db 61 VTDLTHRESDEFSROGNTFLAELGICEPSPHRSYCSDMGILHOGYSLSGSDADSDE 120
Qy 121 GMSPEHATRLNGRGTKSRSSGLSSRENSALTLTDSNENKSDDBNGRPIPTSPSLL 180
Db 121 GMSPEHATRLNGRGTKSRSSGLSSRENSALTLTDSNENKSDDBNGRPIPTSPSLL 180
Qy 181 PSAQLPSSNPPVSCOMPLLDNSNTSHQIMDTNPDEERSPNVYLLRACSGPOQASSGPP 240
Db 181 PSAQLPSSNPPVSCOMPLLDNSNTSHQIMDTNPDEERSPNVYLLRACSGPOQASSGPP 240
Qy 241 NHHSOSTLRPLPPPHNHHTLSHHSSANSLSNLSLNRSSQIHAPAPAPNDLATTPEVQ 300
Db 241 NHHSOSTLRPLPPPHNHHTLSHHSSANSLSNLSLNRSSQIHAPAPAPNDLATTPEVQ 300
Qy 301 LQDSWVLNSVPLETRHFLFKTSSTGPTLFSSSGPYPLTSGTVYTPPPRLLPRNTEFRK 360
Db 301 LQDSWVLNSVPLETRHFLFKTSSTGPTLFSSSGPYPLTSGTVYTPPPRLLPRNTEFRK 360
Qy 361 AFKLLKPSKYCSWKCAALSAIAAALLAILLAYFI----- 395
Db 361 AFKLLKPSKYCSWKCAALSAIAAALLAILLAYFIAMHLLGLNWLOLPADGHTFNNGVYT 420
Qy 396 -----VPMWLNSSIDSGEAEVGRVTVQEVPPGVFWRQSIHISQPFUK 439
Db 421 GLPGNDVATVPSSGKVPWMLNSKSSIDSGEAEVGRVTVQEVPPGVFWRQSIHISQPFUK 480
Qy 440 FNISLGKDALFGVYIRRGLPSPSHAOYDFMERLDGKEKSWVESPRRRSIQTLVNEAVF 499
Db 481 FNISLGKDALFGVYIRRGLPSPSHAOYDFMERLDGKEKSWVESPRRRSIQTLVNEAVF 540
Qy 500 VOYLDVGLWHLAFYNDGDKEMVSENTVVLDVSDQPCPRNCHNGECVSGVCHCFPGFLGA 559
Db 541 VOYLDVGLWHLAFYNDGDKEMVSENTVVLDVSDQPCPRNCHNGECVSGVCHCFPGFLGA 600
Qy 560 DCAKAAKPVLCGNGOYSGTKCQYCSGWKGAECVDPVMOQIDPSCGGHSGCIDGNCVCSA 619
Db 601 DCAKAAKPVLCGNGOYSGTKCQYCSGWKGAECVDPVMOQIDPSCGGHSGCIDGNCVCSA 660
Qy 620 GYKGECEVDCLDPTCSSHGVNCECLSPGWGGLNCELARVOCPDQCSGHGTYLPDT 679
Db 661 GYKGECEVDCLDPTCSSHGVNCECLSPGWGGLNCELARVOCPDQCSGHGTYLPDS 720
Qy 680 GLCSDPNMGMPCSV-VCVSDCGTHGVICGACRCEEGWTGAACDQRYCHPRCIEHGTG 739
Db 721 GLCSDPNMGMPCSV-VCVSDCGTHGVICGACRCEEGWTGAACDQRYCHPRCIEHGTG 779
Qy 740 KDGKCECRGNGEHCITIGROTAGTETDCCPDLCNGNGRCTLGQNSWQCVQOTGHRGPC 799
Db 780 KDGKCECRGNGEHCITIGROTAGTETDCCPDLCNGNGRCTLGQNSWQCVQOTGHRGPC 830
Qy 800 NYAMETSCADNKGEDGLVDCLDPCCLQSAQNSLLCRGSRDPLDIIIOCGQTDWPAVK 859
Db 831 NYAMETSCADNKGEDGLVDCLDPCCLQSAQNSLLCRGSRDPLDIIIOCGQTDWPAVK 890
Qy 860 SFYDRIKLAGKSTHIIIPGENPFFNSLSVSLIRGQVVTDTGTPLVGVNVSFYKPKYGT 919
Db 891 SFYDRIKLAGKSTHIIIPGENPFFNSLSVSLIRGQVVTMDGTPLVGVNVSFYKPKYGT 950
Qy 920 ITRQDGTDLIANGASLTFLHFERAPFMSQRTVMLPWNFSFYAMDTLVNKTTEENIPSCD 979
Db 951 ITRQDGTDLIANGASLTFLHFERAPFMSQRTVMLPWNFSFYAMDTLVNKTTEENIPSCD 1010
Qy 980 LSGFVRPDPPIIISSPLSTFFSAAPQNPVPTQVTLHBEIEPLPGSNVKLRYLSSRTAGYK 1039
Db 1011 LSGFVRPDPPIIISSPLSTFFSAAPQNPVPTQVTLHBEIEPLPGSNVKLRYLSSRTAGYK 1070

Qy 1040 SLLKITMTOSTVPLNLIRVHLMVAVEGHFLQKSFQASPNLASTFIWDKTDAYQORVYGLS 1099
Db 1071 SLLKITMTOSTVPLNLIRVHLMVAVEGHFLQKSFQASPNLAYTFIWDKTDAYQORVYGLS 1130
Qy 1100 DAVSVGFYETPCPSLILWEKRTALLQGFELDPNSLGGWSLDKHHILNVKSGILKHGTGE 1159
Db 1131 DAVSVGFYETPCPSLILWEKRTALLQGFELDPNSLGGWSLDKHHILNVKSGILKHGTGE 1190
Qy 1160 NOFLTQOPALITSIMNGRRRISPCSCNGLAEGNKLLAPVALAVGIDGSLVGVGFNYIR 1219
Db 1191 NOFLTQOPALITSIMNGRRRISPCSCNGLAEGNKLLAPVALAVGIDGSLVGVGFNYIR 1250
Qy 1220 RIFPSRNTVTSILELRNKEFKHNNPAHKYYLAVDPVSGSLYVSDTNSRRIYVKSLSGTFK 1279
Db 1251 RIFPSRNTVTSILELRNKEFKHNSPCHYYLAVDPVSGSLYVSDTNSRRIYVKSLSGAK 1310
Qy 1280 DIAGNSEVVAGTGEQCLPFDEARCGDGGKAIDATLMSPRGIAVKNGLMYFYDATMIRKV 1339
Db 1311 DIAGNSEVVAGTGEQCLPFDEARCGDGGKAVDATLMSPRGIAVKNGLMYFYDATMIRKV 1370
Qy 1340 DONGIISTLLGSDNLTAVRPLSCDSSMDVAOVRLWPTDLAVNPMDSLYVLENNVILRI 1399
Db 1371 DONGIISTLLGSDNLTAVRPLSCDSSMDVAOVRLWPTDLAVNPMDSLYVLENNVILRI 1430
Qy 1400 TENHQVSIITAGRPMHCQVPGIDYSLSKLAHSALESASAIASHGTGLYITETDEKKINR 1459
Db 1431 TENHQVSIITAGRPMHCQVPGIDYSLSKLAHSALESASAIASHGTGLYITETDEKKINR 1490
Qy 1460 LRQVTTNGBICLLAGAAASDCCKNDVNCNCSGDDAYATDAILNSPSSSLAVAPDGTIYTA 1519
Db 1491 LRQVTTNGBICLLAGAAASDCCKNDVNCNCSGDDAYATDAILNSPSSSLAVAPDGTIYTA 1550
Qy 1520 DLGNTRIRAVSNKPNVNAFNOYEAAASPGCEQELYYVFNADGIIHOYTSLVTVGLVNFYTS 1579
Db 1551 DLGNTRIRAVSNKPNVNAFNOYEAAASPGCEQELYYVFNADGIIHOYTSLVTVGLVNFYTS 1610
Qy 1580 TDNDVTELDNNGNSLKIIRRDSSGMPRHLLMPDNOIITLTGTNGGLKVVYSTONLELGLM 1639
Db 1611 ADNDVTELDNNGNSLKIIRRDSSGMPRHLLMPDNOIITLTGTNGGLKAVSTONLELGLM 1670
Qy 1640 TYDGTNGTLLATKSDGTGTTFYDYDHEGRLTNVTRPTGVVTSILHREMEKSIITIDIENSR 1699
Db 1671 TYDGTNGTLLATKSDGTGTTFYDYDHEGRLTNVTRPTGVVTSILHREMEKSIITIDIENSR 1730
Qy 1700 DDDVTVTNLSSVEASYTVVQDVQRNSYQCNNGTLRVMYANGMGSIFSEPHVLGTTIT 1759
Db 1731 DDDVTVTNLSSVEASYTVVQDVQRNSYQCNNGTLRVMYANGMAVSPHSEPHVLGTTIT 1790
Qy 1760 PTIGRCNLSLPMENGLNLSIEWRLKEQIKGKVTIFGRKLRVHGRNLLSIDYDORNIITEXI 1819
Db 1791 PTIGRCNLSLPMENGLNLSIEWRLKEQIKGKVTIFGRKLRVHGRNLLSIDYDORNIITEXI 1850
Qy 1820 YDHRKFTLRILYIDOVGRPFLWLPSSGLAAVNVYFENGRLAGLORGAMSERTDIDKQGR 1879
Db 1851 YDHRKFTLRILYIDOVGRPFLWLPSSGLAAVNVYFENGRLAGLORGAMSERTDIDKQGR 1910
Qy 1880 IVSRMFADCKVMSYSLDKSMVLLLSQSQRYIFEDYDSSDRLLAVTMPSVARHSMSTHTSI 1939
Db 1911 IVSRMFADCKVMSYSLDKSMVLLLSQSQRYIFEDYDSSDRLLAVTMPSVARHSMSTHTSI 1970
Qy 1940 GYIRNIYNPPESNASVIEDYSDDGRLKTSFGLGTGRQVYKYGKLSKJSEIYVYDSTAVTF 1999
Db 1971 GYIRNIYNPPESNASVIEDYSDDGRLKTSFGLGTGRQVYKYGKLSKJSEIYVYDSTAVTF 2030
Qy 2000 GYDETGTGVLKMNLOSOGGFSCTIRYRKICPLVDKQIYREFSEEGKVNARFDYTHONSRI 2059
Db 2031 GYDETGTGVLKMNLOSOGGFSCTIRYRKICPLVDKQIYREFSEEGKVNARFDYTHONSRI 2090
Qy 2060 ASIKPVISETPLVDLYRYDEISGKVEHFGKFGVYIYDINOIITAVMTLSKHFTDHGRI 2119
Db 2091 ASIKPVISETPLVDLYRYDEISGKVEHFGKFGVYIYDINOIITAVMTLSKHFTDHGRI 2150

Qy 2120 KEVYEMFRSLMYWMTVOYDSMGVRVIRKELKGLGYANTTKYTYDYGQGLQSVAVVNDP 2179
Db 2151 KEVYEMFRSLMYWMTVOYDSMGVRVIRKELKGLGYANTTKYTYDYGQGLQSVAVVNDP 2210
Qy 2180 TWRSYDLNGLNHLNPGNSVRLMPLRYDRLDRITRGLDVOYKIDDDGYLCORGSDIFEY 2239
Db 2211 TWRSYDLNGLNHLNPGNSARLMLPLRYDRLDRITRGLDVOYKIDDDGYLCORGSDIFEY 2270
Qy 2240 NSKGLLFRAYNKASGWSQVRYDCVGRASYKTNLGHHLQYFYSDLNHPNTRITHVYNSN 2299
Db 2271 NSKGLLFRAYNKASGWSQVRYDCVGRASYKTNLGHHLQYFYSDLNHPNTRITHVYNSN 2330
Qy 2300 SEITSLYDLOGLHFLAFEMESSGEYEVASDNTGTPPLAVFSINGLMIKOLOQYATYGEIYD 2359
Db 2331 SEITSLYDLOGLHFLAFEMESSGEYEVASDNTGTPPLAVFSINGLMIKOLOQYATYGEIYD 2390
Qy 2360 SNPDFQMVICPHGLYDPLTKLVHFTQRDYDVLGRWTSPTYTMWKNVKGEPAPFNLYMF 2419
Db 2391 SNPDFQMVICPHGLYDPLTKLVHFTQRDYDVLGRWTSPTYTMWKNVKGEPAPFNLYMF 2450
Qy 2420 KSNPLSSELDLKNVYDVKSWMVFGQLSNIIIPGPRAKMYFVPPPYELSESQASENG 2479
Db 2451 KSNPLSSELDLKNVYDVKSWMVFGQLSNIIIPGPRAKMYFVPPPYELSESQASENG 2510
Qy 2480 QLITGVQOOTTERHNOAFMALEGOVITTKLHASIREKAGHWFATTPPIIGKIMFAIKR 2539
Db 2511 QLITGVQOOTTERHNOAFMALEGOVITTKLHASIREKAGHWFATTPPIIGKIMFAIKR 2570
Qy 2540 VTTGVSSITASDSRKVASLVNNAYLDKMHYSIEGKOTHYFVKIGSADGDLVLTGTTGR 2599
Db 2571 VTTGVSSITASDSRKVASLVNNAYLDKMHYSIEGKOTHYFVKIGSADGDLVLTGTTGR 2630
Qy 2600 KVLSEGVNVTVSQPTLLVNGTRFTNIEFOYSTLLSIRYGLTPDTPDEKARVLQAR 2659
Db 2631 KVLSEGVNVTVSQPTLLVNGTRFTNIEFOYSTLLSIRYGLTPDTPDEKARVLQAR 2690
Qy 2660 QRALGTAWAKEQQKARDGREGSLRWTEGEKQOLLSTGRVQYEGYVYLPVQYPELADSS 2719
Db 2691 QRALGTAWAKEQQKARDGREGSLRWTEGEKQOLLSTGRVQYEGYVYLPVQYPELADSS 2750
Qy 2720 SNIQFLQKQNGMKR 2733
Db 2751 SNIQFLQKQNGMKR 2764

RESULT 2
Q9R1K2 PRELIMINARY; PRT: 2765 AA.
Q9R1K2; 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE NEURESTIN ALPHA.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE DAWLEY; TISSUE=OLFACTORY BULB;
RX MEDLINE=93350226; PubMed=10419693;
RA Ozaki J.M., Firestein S.;
RT "Neurestin: putative transmembrane molecule implicated in neuronal
development.";
RL Dev. Biol. 212:165-181(1999).
DR EMBL: AF086607; AAD47393.1; -
DR HSSP: P00750; ITPG.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR002049; Laminin_EGF.
DR InterPro: IPR001258; NHL.
DR Pfam: PF000008; EGF; 4.
DR Pfam: PF01436; NHL; 2.
DR PRINTS: PR00011; EGF/LAMININ.

DR SMART; SMO0181; EGF; 5.
DR PROSITE; PS00022; EGF_1; UNKNOWN_8.
DR PROSITE; PS01186; EGF_2; 7.
KW EGF-like domain; Glycoprotein.
SQ SEQUENCE 2765 AA; 306568 MW; 6748D70D5FFD8F0E CRC64;

Query Match 98.0%; Score 14291; DB 11; Length 2765;
Best Local Similarity 96.5%; Pred. No. 0;
Matches 2676; Conservative 30; Mismatches 18; Indels 50; Gaps 2;

Qy 1 MDVKDRHRSLTRGRCCKEYRTSSSLDSEDCRVPTOKSYSSSETLKAYDHDHSMHYGNR 60
Db 1 MDVKDRHRSLTRGRCCKEYRTSSSLDSEDCRVPTOKSYSSSETLKAYDHDHSMHYGNR 60
Qy 61 VTDLIHRESDEFPRQGFNFTLAEIGICEPSPHRSGYCDMGLHGGYSLSTGSDADSDTE 120
Db 61 VTDLIHRESDEFPRQGFNFTLAEIGICEPSPHRSGYCDMGLHGGYSLSTGSDADSDTE 120
Qy 121 GMSPEHAIRLWGRGKSRSSGLSSRENSALTLTDSNENKSDDENGRPPTSSPSSL 180
Db 121 GMSPEHAIRLWGRGKSRSSGLSSRENSALTLTDSNENKSDDENGRPPTSSPSSL 180
Qy 181 PSAQLPSSHNPVVSQMPPLDLSNTHQIMDTNPDEEFSPNSYLLRACSGPQQASSGPP 240
Db 181 PSAQLPSSHNPVVSQMPPLDLSNTHQIMDTNPDEEFSPNSYLLRACSGPQQASSGPP 240
Qy 241 NHHQSOTLRPLPPPHNHTLSHHSSANSLSNRSLTNRRSOIHAPAPAPNDLATTPEVQ 300
Db 241 NHHQSOTLRPLPPPHNHTLSHHSSANSLSNRSLTNRRSOIHAPAPAPNDLATTPEVQ 300
Qy 301 LQDSWLVNSNVPLETRHFLFKTSSGSTPLFSSSSPGYPLTSGTYVTPPRLLPNTFSRK 360
Db 301 LQDSWLVNSNVPLETRHFLFKTSSGSTPLFSSSSPGYPLTSGTYVTPPRLLPNTFSRK 360
Qy 361 AFKLKSKYKSWKCAALSATAALLAILLAYFI-----VPWSLKSSIDSGEAEVGRRTVQEVPPGVFWRSSQTHISQPF 395
Db 361 AFKLKSKYKSWKCAALSATAALLAILLAYFIAMHLLGLNQLQPADGHTFNNGVRT 420
Qy 396 -----VPWSLKSSIDSGEAEVGRRTVQEVPPGVFWRSSQTHISQPF 439
Db 421 GLPGNDVAIVPSGGKVPWSLKSSIDSGEAEVGRRTVQEVPPGVFWRSSQTHISQPF 480
Qy 440 FNIISLGDALFVYVIRRLPSPHAQYDFMERLDGKEXWVSVESPRRRSIQTLVQNEAVF 499
Db 481 FNIISLGDALFVYVIRRLPSPHAQYDFMERLDGKEXWVSVESPRRRSIQTLVQNEAVF 540
Qy 500 VOYLDVGLWHLAFYNDGKDKEMVSFNTVVLDSVQDCPNCHGNGECVSGVCHCFPGFLGA 559
Db 541 VOYLDVGLWHLAFYNDGKDKEMVSFNTVVLDSVQDCPNCHGNGECVSGVCHCFPGFLGA 600
Qy 560 DCAKAACPVLCSGNGQYSGTCQCYSGWKGAECVPMNQCIDPSCGGHSGCIDGNCVCSA 619
Db 601 DCAKAACPVLCSGNGQYSGTCQCYSGWKGAECVPMNQCIDPSCGGHSGCIDGNCVCSA 660
Qy 620 GYKGEHCEVDCLDPTCSSHGCVNGECLSPGNGGLNCELARVQCPCQCSGHGTYLPDT 679
Db 661 GYKGEHCEVDCLDPTCSSHGCVNGECLSPGNGGLNCELARVQCPCQCSGHGTYLPDT 720
Qy 680 GLCSCDPNMGWPCDSVEVCSVDCGTHGVCIGGACRCEBGTGAACDQVRVCHPRCIEHCTC 739
Db 721 GLCSCDPNMGWPCDSVEVCSVDCGTHGVCIGGACRCEBGTGAACDQVRVCHPRCIEHCTC 780
Qy 740 KDGKCECEGNGEHCITIGRTAGTETDGCPCDLNCHGRCCTLGNSWQCVQCTQWRGPGC 799
Db 781 KDGKCECEGNGEHCITIGRTAGTETDGCPCDLNCHGRCCTLGNSWQCVQCTQWRGPGC 831
Qy 800 NVAMETSCADKNKDEGLVDCLDPCCLQSAQNSLLCRGSRDPLDIIOQQTDPWPAVK 859
Db 832 NVAMETSCADKNKDEGLVDCLDPCCLQSAQNSLLCRGSRDPLDIIOQQTDPWPAVK 891
Qy 860 SFYDRILKAGKDSHTIIPGENPNSLSLIRGVVTTDGTPLGVNVSFVKPKYGYT 919
Db 919 SFYDRILKAGKDSHTIIPGENPNSLSLIRGVVTTDGTPLGVNVSFVKPKYGYT 919

Db 892 SFYDRIKLAGKSTHIIIPGDNFNSLSLIRGQVVTGDTGLVGVNVSFKYKPYGT 951
Qy 920 ITRQDGTFLIANGASLTLHFERAPFMQERTVLPWNSFYAMDPLVMKTEENSPSCD 979
Db 952 ITRQDGTFLIANGASLTLHFERAPFMQERTVLPWNSFYAMDPLVMKTEENSPSCD 1011
Qy 980 LSGFVRPDPPIIISSPLSTFFSAAPQNPVTPQVQLHHEIELPGSNVKLYLSSRTAGYK 1039
Db 1012 LSGFVRPDPPIIISSPLSTFFSAAPNPVTPQVQLHHEIELPGTNVKLYLSSRTAGYK 1071
Qy 1040 SLLKTIWTOSTVPLNIRVHLVAVVEGHLFQKSFQASPNLSTFFIWDKTDAYQORVYGLS 1099
Db 1072 SLLKTIWTOSTVPLNIRVHLVAVVEGHLFQKSFQASPNLAYTFFIWDKTDAYQORVYGLS 1131
Qy 1100 DAVSVGFYEYETCPSLILWEKRTALLOQFELDPNSLGGNSLKHHTLNKSGILKHGTGE 1159
Db 1132 DAVSVGFYEYETCPSLILWEKRTALLOQFELDPNSLGGNSLKHHTLNKSGILKHGTGE 1191
Qy 1160 NOFLTQOQPAIITSINGNRRRRISPCSCNGLAEGNKLAPALAVIGDGLSYGVDFNYIR 1219
Db 1192 NOFLTQOQPAIITSINGNRRRRISPCSCNGLAEGNKLAPALAVIGDGLSYGVDFNYIR 1251
Qy 1220 RIFFPSRNTSILELRNKEPKHNNPAHKYVYLAVDPVSGSLYVSDTNSRRIRYRVKSLSGTK 1279
Db 1252 RIFFPSRNTSILELRNKEPKHNSPGHKYVYLAVDPVGTGSLYVSDTNSRRIRYRVKSLSGAK 1311
Qy 1280 DLAGNSEVVAGTGEQCLPFDEARCGDGGKAIDATLMSPRGIAVDKNGLMYFVDATMIRKV 1339
Db 1312 DLAGNSEVVAGTGEQCLPFDEARCGDGGKAVDATLMSPRGIAVDKNGLMYFVDATMIRKV 1371
Qy 1340 DQNGIISTLLGNDLTAVRPLSCDSSMDVAQVRLWPTDLAVNPMDNSLYVLENNVILRI 1399
Db 1372 DQNGIISTLLGNDLTAVRPLSCDSSMDVAQVRLWPTDLAVNPMDNSLYVLENNVILRI 1431
Qy 1400 TENHQVSIITAGPMHCQVPGIDYLSKLAIHSALESASAIATSHGTGLYITETDEKKINR 1459
Db 1432 TENHQVSIITAGPMHCQVPGIDYLSKLAIHSALESASAIATSHGTGLYITETDEKKINR 1491
Qy 1460 LRQVTTNGEICLLAGAAASDCCKNDVNCYSGDDAYATDAILNSPSSLVAVPDGTIYIA 1519
Db 1492 LRQVTTNGEICLLAGAAASDCCKNDVNCICYSGDDAYATDAILNSPSSLVAVPDGTIYIA 1551
Qy 1520 DLGNTRIRAVSNKPKVLNAFNQYEAASPGQEELYVFNADGIIHQYTVSLVTGEYLYNFTYS 1579
Db 1552 DLGNTRIRAVSNKPKVLNAFNQYEAASPGQEELYVFNADGIIHQYTVSLVTGEYLYNFTYS 1611
Qy 1580 TDNDVTELDNNGNSLKIIRDSGMPHRLIMPNDQIITLTGVTNGGLKVWSTONLELGLM 1639
Db 1612 ADNDVTELDNNGNSLKIIRDSGMPHRLIMPNDQIITLTGVTNGGLKAVSTONLELGLM 1671
Qy 1640 TYDGNLTGLLATKSDEGTWTFYDYDHEGRLTNVTRPTGVVTSLSHREMEKSIITIDIENSR 1699
Db 1672 TYDGNLTGLLATKSDEGTWTFYDYDHEGRLTNVTRPTGVVTSLSHREMEKSIITIDIENSR 1731
Qy 1700 DDDVTVTNLSSEASVTWQOVNRNSYQLCNNGTLRVNANGMGSIFHSEPHVLAGTIT 1759
Db 1732 DNDVTVTNLSSEASVTWQOVNRNSYQLCNNGTLRVNANGMGSIFHSEPHVLAGTIT 1791
Qy 1760 PTIGRCNISLPMENGLNTEWRLRKEQIKGVTIFGRKLVRHGRNLLSIDYDRNIRTEKI 1819
Db 1792 PTIGRCNISLPMENGLNTEWRLRKEQIKGVTIFGRKLVRHGRNLLSIDYDRNIRTEKI 1851
Qy 1820 YDHRKFTLRIIYDQVGRPFLWLPSSGLAAVNVSYFFNGRLAGLQRGAMSERDIDKQGR 1879
Db 1852 YDHRKFTLRIIYDQVGRPFLWLPSSGLAAVNVSYFFNGRLAGLQRGAMSERDIDKQGR 1911
Qy 1880 IYSRMPADGKVMWSYLDKSWLLQLSQORQYIFEYDSSDRLLAVTMPVSARHSMSTHTSI 1939
Db 1912 IYSRMPADGKVMWSYLDKSWLLQLSQORQYIFEYDSSDRLLAVTMPVSARHSMSTHTSI 1971
Qy 1940 GYIRINYNPESNASVIFDYSDGRILKTFLTGCRQVFKYKGLSKLSKEIYVDSTAVTF 1999
Db 1972 GYIRINYNPESNASVIFDYSDGRILKTFLTGCRQVFKYKGLSKLSKEIYVDSTAVTF 2031

Qy 2000 GYDETTGVLKMYNLQSGGFSCTIRYRKIGPLVDKQIYREFSEGMVNAFDTYTHNSFRI 2059
Db 2032 GYDETTGVLKMYNLQSGGFSCTIRYRKIGPLVDKQIYREFSEGMVNAFDTYTHNSFRI 2091
Qy 2060 ASIKPVISTPLVDLYRYDEISGKVEHFGKFGVYIYDINOIITAVMTLSKHFTHGRI 2119
Db 2092 ASIKPVISTPLVDLYRYDEISGKVEHFGKFGVYIYDINOIITAVMTLSKHFTHGRI 2151
Qy 2120 KEVOYEMERSLWMTVQYDSMGRVTKRELKLGPAVNTTKYTDYDGGLOLOSVAVNDP 2179
Db 2152 KEVOYEMERSLWMTVQYDSMGRVTKRELKLGPAVNTTKYTDYDGGLOLOSVAVNDP 2211
Qy 2180 TWRYSYDLNGLNHLNPGNSVRLMPLRYDLRDRITRLGDOVQYKIDDOGYLCOGSDIFEY 2239
Db 2212 TWRYSYDLNGLNHLNPGNSVRLMPLRYDLRDRITRLGDOVQYKIDDOGYLCOGSDIFEY 2271
Qy 2240 NSKGLLTRYANKASGWSVOYRYDGVGRASRYKTNLGHLOFYFSDLNHPTRTHYVNSN 2299
Db 2272 NSKGLLTRYANKASGWSVOYRYDGVGRASRYKTNLGHLOFYFSDLNHPTRTHYVNSN 2331
Qy 2300 SEITSLYDLQGLHFLAMESSEGEYVYASDNTGTPLAVFSINGLMIKOLQYTAIGEYVD 2359
Db 2332 SEITSLYDLQGLHFLAMESSEGEYVYASDNTGTPLAVFSINGLMIKOLQYTAIGEYVD 2391
Qy 2360 SNPDFQMVIGFHGGLYDPLTKLVHFTQDYOVLAGRWTSPOYTMKNVKGEPAPFNLYM 2419
Db 2392 SNPDFQMVIGFHGGLYDPLTKLVHFTQDYOVLAGRWTSPOYTMKNVKGEPAPFNLYM 2451
Qy 2420 KSNPLSELDLKNVYTDVKSMLVMFGQLSNIIIPGFPRAKMYFVPPPELSESOASENG 2479
Db 2452 KSNPLSELDLKNVYTDVKSMLVMFGQLSNIIIPGFPRAKMYFVPPPELSESOASENG 2511
Qy 2480 QUITGVQOOTERRHNOAFMALEGOVITTKLHASIREKAGHWFATTPIIGKGMFAIKER 2539
Db 2512 QUITGVQOOTERRHNOAFMALEGOVITTKLHASIREKAGHWFATTPIIGKGMFAIKER 2571
Qy 2540 VTTGVSSIASDSRKSASVNLNAYYLDKMHYSIEGKTDHYFVKISAGDGLVTLGTTIGR 2599
Db 2572 VTTGVSSIASDSRKSASVNLNAYYLDKMHYSIEGKTDHYFVKISAGDGLVTLGTTIGR 2631
Qy 2600 KYLESVNVTVSQPTLLVNGRTRRTNIEFOYSTLLSTRYGLTPTDLDDEKARVLDQAR 2659
Db 2632 KYLESVNVTVSQPTLLVNGRTRRTNIEFOYSTLLSTRYGLTPTDLDDEKARVLDQAR 2691
Qy 2660 QALGTAWAKEQOKARDREGSRLWTEGEKQOLLSTGRVQGYEYVLPVEQYPELADSS 2719
Db 2692 QALGTAWAKEQOKARDREGSRLWTEGEKQOLLSTGRVQGYEYVLPVEQYPELADSS 2751
Qy 2720 SNIQFLRONEMGR 2733
Db 2752 SNIQFLRONEMGR 2765

RESULT 3

Q5DERS
ID Q5DERS PRELIMINARY; PRT; 2802 AA.
AC Q5DERS;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE TENEURIN-2 (FRAGMENT).
GN TEN2.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN
RP SEQUENCE FROM N.A.
RX MEDLINE=20581705; PubMed=11146505;
RA Tucker R.P., Chiquet-Ehrismann R., Chevrion M., Martin D., Hall R.J.,
RA Rubin B.P.;

RT	"Teneurin-2 is expressed in tissues that regulate limb and somite									
RL	pattern formation and is induced in vitro and in situ by FGF8.,"									
DR	EMBL; AJ279031; CAC09416.1; --									
DR	HSBP; P00750.1; ITPG.									
DR	InterPro; IPR000561; EGF-like.									
DR	InterPro; IPR002049; Laminin_EGF.									
DR	Pfam; PF00008; EGF; 5.									
DR	PRINTS; PR00011; EGF_LAMININ.									
DR	SMART; SM00181; EGF; 7.									
DR	PROSITE; PS00022; EGF_1; UNKNOWN_1.									
FT	NON_TER	2802	2802	1						
SQ	SEQUENCE	2802	AA; 310745	MW; B1FBC2C84EDFA4B3	CRC64;					
<hr/>										
Query Match 95.68; Score 13934.5; DB 13; Length 2802;										
Best Local Similarity 92.38; Pred. No. 0;										
atches 2586; Conservative 95; Mismatches 52; Indels 69; Gaps 2;										
QY	1	MDVKRRHRS	LTGRGCKRECRYTSSSLDSEDCRVPTOKSYSSSETLKAYDHSRMHYGNR	60						
DB	1	MDIKDRHRS	LTGRGCKRECRYTSSSLDSEDCRVPAQKSYSSSETLKAYGHDTRMHYGNR	60						
QY	61	VTDLTHRSDE	FPROGNTNFTLAELGICEPSPHRSGYCDMGILHOGYSLSTGSDADSDE	120						
DB	61	VSDLVHRSDE	FPROGNTNFTLAELGICEPSPHRSGYCDIGILHOGYSLSTGSDADSDE	120						
QY	121	GMSPHAIRL	WGRGKISRRSGLSRENSALTLTDSNENKSDENG-----	168						
DB	121	GMSPHAIRL	WGRGKISRRSGLSRENSALTLTDSNENKSDENDFHTLSKLDKDR	180						
QY	169	-----	RRIPPTSSPLPSAQLPSSHNPVPVSCOMPLLDNSTHQIMDT	212						
DB	181	QTSWQOLA	ETKNSLRIRPTSSSLLPSAQLPSSHNPVPVSCOMPLLDNSTHQIMDT	240						
QY	213	NPDEFSN	SYLLRACSGPQQAASSGPPNHHQSQTLRPLPPPHNHTLSHHSSANSNR	272						
DB	241	NPDEFSN	SYLLRACSGPQQAASSGPPNHHQSQTLRPLPPPHNHTLSHHSSANSNR	300						
QY	273	NSLTNRQI	HAPAPANDLATTPEVOLQDSWLNWNPVLETRHFLFKTSGGSTPLFSS	332						
DB	301	NSLTNRQI	HAPAPANDLATTPEVOLQDSWLNWNPVLETRHFLFKTSGGSTPLFSS	360						
QY	333	SSPGYPLT	SGTYVTPPPRLPNTESRKAFLKPKSKYCSWKCAALSAIAAALLAILLA	392						
DB	361	SSPGYPLT	SGTYVTPPPRLPNTESRKAFLKPKSKYCSWKCAALSAIAAALLAILLA	420						
QY	393	YPIV-----	-----PWSLKNSSIDSGAE	411						
DB	421	YPIAMHLL	GLNLQNPADGHTFSNGLRPGAAGAEAGAAAPAGRGPVWTRNSSIDSGETE	480						
QY	412	VGRRYTQ	VEPPGVFWRSOIHSOPQFLKFNISLGKDALFGVYIRGLPPSHAQYDFMERL	471						
DB	481	VGRKYTQ	VEPPGVFWRSOIHSOPQFLKFNISLGKDALFGVYIRGLPPSHAQYDFMERL	540						
QY	472	DGKEKWS	VVESPRERSIQTLVQNEAVFYQYLDVGLWHLAFYNDGKDEKEMYSFNTVWLDS	531						
DB	541	DGKEKWS	VVESPRERSIQTLVQNEAVFYQYLDVGLWHLAFYNDGKDEKVEVSFNTVILDS	600						
QY	532	VODCPRN	CHNGECVSGVCHCFPGFLGADCAKAAAPVLCGSGNGOYSKGTCCOYSGWKGA	591						
DB	601	VODCPRN	CHNGECVSGVCHCFPGFLGADCAKAAAPVLCGSGNGOYSKGTCLCYSGWKGE	660						
QY	592	CDVPMNO	CIIDPSCGSGCIDGNCVCSAGYGEHCHEEVDCLDPTCSSHGVCVNGECLCSP	651						
DB	661	CDVPI	ISQCIIDPSCGSGSGIEGNCVCSIGYKGENCEEVDCLDPTCSNHGVCVNGECLCSP	720						
QY	652	GWGLNCEL	ARVQCPCQSGHGYTLPDTGLCSDCPNMGPDGSEVVCVSDCGTHGVCIGG	711						
DB	721	GWGINCEL	PARQPCQSGHGYTLPDTGLCSDCPNMGPDGSEVVCVSDCGTHGVCIGG	780						
QY	712	ACRCEEG	WTGAACDQRVCHPRCIEHGTCKDKGKCECREGNGEHCTIGROTAGETDGC	771						

DB	781	ACRCEEG	WTGACDQRVCHPRCTEHTGCTCKDKGKCECREGNGEHCTIGROTAGETDGC	840						
QY	772	LCNGNGR	CTLGQNSWQVCQGTGWRGPGCNVAMETSCADKNKNEGDLVCDLDPCCLOSA	831						
DB	841	LCNGNGR	CTLGQNSWQVCQGTGWRGPGCNVAMETSCADKNKNEGDLVCDLDPCCLOST	900						
QY	832	QNSLLCR	SGRDLPIIQOQGTDMPAKVSFYDRIKLLAGKDSSTHIIIPGENFNSSLSVLI	891						
DB	901	QNSLLCR	SGRDLPIIQOHSGPSAVASFYDRIKLLVKGKDSSTHIIIPGENFNSSLSVLI	960						
QY	892	RQVVYTD	GTPLVGVNVSFYKPKYKTYITRODGTFDLIANGASLTUHFERAPMSQER	951						
DB	961	RQVVYTD	GTPLVGVNVSFYKPKYKTYITRODGMFVLVANGSSLTUHFERAPMSQER	1020						
QY	952	TVLWPNS	FYAMDTLVMKTEENSIPSCDLSGFVRDPDIIISPLSTFFSAAPGNPIYPE	1011						
DB	1021	TVLWPNS	FYAMDTLVMKTEENSIPSCDLSGFVRDPDIIISPLSTFFSDAPGRNPIYPE	1080						
QY	1012	TQVLHEE	TELPGSKVRLYSRTAGYKSLKIKITMTQSTVPLNLRVHLMAVEGHLFQK	1071						
DB	1081	TQVLHEE	TEVPSSIKLIYLSRTAGYKSLKIKITMTQSLVPLNLKVLHMAVEGHLFQK	1140						
QY	1072	SFOASPN	LASTFINDKDAYGORVYGLSDAVVSVGFYETCPSLLILWEKRTALLOGFELD	1131						
DB	1141	SFLASPN	LAYTFINDKDAYGQVYGLSDAVVSVGFYETCPSLLILWEKRTALLOGFELD	1200						
QY	1132	PSNLGGW	SLDKHHILNVKSGILHKGTGENOFLOTPAIIITSIMGNRRRSISCPSCNGLA	1191						
DB	1201	PSNLGGW	SLDKHHILNVKSGILHKGTGENOFLOTPQAVITSIMGNRRRSISCPSCNGLA	1260						
QY	1192	EGNKLLA	PVALAVAGTDGSLYVGDFNYIRRIIPSRNVTSILELRNKEFKHNNPAHKYLA	1251						
DB	1261	EGNKLLA	PVALAVAGTDGSLYVGDFNYIRRIIPSRNVTSILELRNKEFKHNNPAHKYLA	1320						
QY	1252	VDPVSGS	LYVSDTNSRRIRYRVKSLGTDKDLAGNSEVAVAGTGEQCLPFEARCGDGGKAID	1311						
DB	1321	VDPVSGS	LYVSDTNSRRIRYRVKSLGTDKDLAGNSEVAVAGTGEQCLPFEARCGDGGKAVD	1380						
QY	1312	ATLMSPR	GIADVKNGLMYFVDATMIRKVDQNGIISTLGSNDLTAVRPLSCDSSMDVAQV	1371						
DB	1381	ATLMSPR	GIADVKNGLMYFVDATMIRKVDQNGIISTLGSNDLTAVRPLSCDSSMDVQV	1440						
QY	1372	RLEWPTD	LAVAMPDMSLYVLNNVILRTENHQVSIAGRPMHCQVPGIDYSLSLAHS	1431						
DB	1441	RLEWPTD	LAVAMPDMSLYVLNNVILRTENHQVSIAGRPMHCQVPGIDYSLSLAHS	1500						
QY	1432	ALESASA	IAISHTGVLYITETDEKKINRLQVTTNGEICLLAGAAASDCCKNDVNCNYS	1491						
DB	1501	ALESASA	IAISHTGVLYISETDEKKINRLQVTTNGEICLLAGAAASDCCKNDVNCNYS	1560						
QY	1492	GDDAVAT	AILNPSLAVAPDGTIYIADLGNIRIRAVSKNKPVLNAPNOYEAAASPGQE	1551						
DB	1561	GDDGAT	AILNPSLAVAPDGTIYIADLGNIRIRAVSKNKPVLNAPNOYEAAASPGQE	1620						
QY	1552	LYVFNAD	GHIQVYTSVLTGEYLYNFTYTDNDVTDLIDNNGNSLIRRDSSGMPRHLLMP	1611						
DB	1621	LYVFNAD	GHIQVYTSVLTGEYLYNFTYSSDNDVTEDVMSNGNSLAVRRDASGMPRHLLMP	1680						
QY	1612	DNQIITL	VJGVNGGLKVYSTQNLEGLMTYDGTGTLATKSDGTGTTTFYDHDHGRLTN	1671						
DB	1681	DNQIYTL	AVGTNGGLKVYSTQLEGLMTYNGNSGLLATKSDGTGTTTFYDHDHGRLTN	1740						
QY	1672	VTRPTGV	VVTSLHREMEKSTTIDDIENSNRDDVTITNLSNVSEASVTVVQDQVRNSYQLCN	1731						
DB	1741	VTRPTGV	VVTSLHREMEKSTTIDDIENSNRDDVTITNLSNVSEASVTVVQDQVRNSYQLCN	1800						
QY	1732	NGTLRMV	ANGMISFHPSEPHVLAQITPTIGRCNISLPMENGLNSIEWRLRKEQIKGV	1791						
DB	1801	NGTLRMV	ANGMISFHPSEPHVLAQITPTIGRCNISLPMENGLNSIEWRLRKEQIKGV	1860						
QY	1792	TIFGRKL	RVHGRNLLSIDYDRNIRTEKIYDDHRKFTLRITIIYDQVGRPFLWLPSSGLA	1851						

Db	1861	TVFGRKLVRHGRNLLSIDYDRNRTRTEKIIDDRHKFTLRIIYDQLGPRFLWLPSSGLAAVN	1920
QY	1852	VSFFNGRLAGLQRGAMSERTDIDKGRIVSRMFADGKVMYSYGLDKSMVLLLSQSORQYI	1911
Db	1921	VSFFNGRLAGLQRGAMSERTDIDKGRIFSRMFADGKVMYSYILEKSMVLLLSQSORQYI	1980
QY	1912	FYDSSDRLLAVTMPSPVARHSMSTHTSIGYIRNIYNPPESNASVIFDYSDDGRILKTSFL	1971
Db	1981	FYDSSDRLLHAVTMPSPVARHSMSTHTSVGIRNIYNPPESNASVIFDYSDDGRILKTSFL	2040
QY	1972	GTCRQVFFYKGLKSLSEIYVDSTAVTFGVDYTGVLKWNVLQSGGFSCTIRYRKIGPLV	2031
Db	2041	GTCRQVFFYKGLKSLSEIYVDSTAVTFGVDYTGVLKWNVLQSGGFSCTIRYRKIGPLV	2100
QY	2032	DKOYRFSEGMVNARFDVTHDNSFRIASIKPVISETPLPVDLYRYDEISGKVEHFGKF	2091
Db	2101	DKOYRFSEGMVNARFDVTHDNSFRIASIKPIISETPLPVDLYRYDEISGKVEHFGKF	2160
QY	2092	GVIIYDINQIITAVMTLSKHFDTHGRIRKEVQYEMFRSLMYWMTVOYDSMGRVYKRELK	2151
Db	2161	GVIIYDINQIITAVMTLSKHFDTHGRIRKEVQYEMFRSLMYWMTVOYDSMGRVYKRELK	2220
Db	2152	GPYANTTKTYDYGQQLQSVAVNDPRTWRYSYDNLGNLHLLNPGNSVRLMPLRYDLRD	2211
Db	2221	GPYANTTKTYDYGQQLQSVAVNDPRTWRYSYDNLGNLHLLNPGNSVRLMPLRYDLRD	2280
QY	2212	RITRLGDVOYKIDDDGYLQCGSDIFEYNSKGLLTRYANKASGWSYQYRDYGVGRASYK	2271
Db	2281	RITRLGDIYKIDDDGYLQCGSDIFEYNSKGLLTRYANKANGWVQYRYDGLGRASCK	2340
QY	2272	TNLGHHLQFYSDLNHPTRITHVYHNSSEITSLYDLOGLHLPAMSSSGEYEVVADSN	2331
Db	2341	TNLGHHLQFYADLNHPTRVTHVYHNSSEITSLYDLOGLHLPAMSSSGEYEVVADSN	2400
QY	2332	GTPPLAVFSLNGLMIKOLQYATGEIYYDSNPDPQWIGVHGGGLYDPLTKLVHFTORDYD	2391
Db	2401	GTPPLAVFSLNGLMIKOLQYATGEIYYDSNPDPQWIGVHGGGLYDPLTKLVHFTORDYD	2460
QY	2392	LGRWTSPOYTMKNYKGEKAPENLNFYKSNPNLSSELDELKNYVTDVKSWMVFGFQLSN	2451
Db	2461	LGRWTSPOYTMKNYKGEKAPENLNFYKSNPNLSNELDLKNYVTDVKSWMVFGFQLSN	2520
QY	2452	IIPGFRKMYFVPPPYELSESAQSENGQLITGVQOTTERRHNOAFMALEGQVITKKLHAS	2511
Db	2521	IIPGFRKMYFVPPPYELTESQACENGQLITGVQOTTERRHNOAFMALEGQVLSKRLHAS	2580
QY	2512	IREKAGHWFATTPITIGKIMFAIKGGRVTVGVSSIASDESDSRKVASVLNNAYYLDKMHYS	2571
Db	2581	IREKAGHWFATSTPIIGKIMFAVKGRTVTGISSIATDDSRKIASVLSNAHYLEKMHYS	2640
QY	2572	IEGKDTYFVKIGSADGDLVTGTTIGRKVLESQVNVTVSQPTLLVNGRTRRTNIEFOY	2631
Db	2641	IEGKDTYFVKIGSADSLVTLAMTSGRKVLDSGVNVTVSQPTLLNGRTRRTNIEFOY	2700
QY	2632	STLLLSIRYGLPTDLDDEKARVLDQARALGTAWAKEQKARDREGSRWLWTEGEKQ	2691
Db	2701	STLLINIRYGLTADTLDDEKARVLDQARALGSAAWAKEQKARDREGSRVWTDGEKQ	2760
QY	2692	LLSTGRVQVGEYVLPVEQYPELADSSNIOFLRONEMGKR 2733	
Db	2761	LLNTGRVQVGEYVLPVEQYPELADSSNIOFLRONEMGKR 2802	
RESULT	4		
Q9WTS6			
ID	Q9WTS6	PRELIMINARY;	PRT; 2715 AA.
AC	Q9WTS6;		
DT	01-NOV-1999 (TremBLrel. 12, Created)		
DT	01-NOV-1999 (TremBLrel. 12, Last sequence update)		
DT	01-DEC-2001 (TremBLrel. 19, Last annotation update)		
DE	TEN-M3.		
GN	OD23 OR TEN-M3.		
OS	Mus musculus (Mouse).		

OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Mammalia; Rodentia; Sciurognathi; Muridae; Mus.		
OX	NCBI_TaxID=10090;		
RN	(1)		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=BALB/C; TISSUE=BRAIN;		
RA	Oohashi T., Zhou X., Feng K., Richter B., Moergelin M., Perez M.T.,		
RA	Su W., Chiquet-Ehrismann R., Rauch U., Faessler R.;		
RT	"Mouse Ten-m/Odz is a new family of dimeric type II transmembrane		
RT	proteins expressed in many tissues.";		
RL	J. Cell Biol. 0-0-0(1999).		
DR	EMBL; AB025412; BAA77398.1; -.		
DR	HSSP; P01452; IVNB.		
DR	MGD; MGI:1345183; Odz3.		
DR	InterPro; IPR000561; EGF-like.		
DR	InterPro; IPR002049; Laminin_EGF.		
DR	Pfam; PF00008; EGF_6.		
DR	PRINTS; PR00011; EGFLAMININ.		
DR	SMART; SM00181; EGF_6.		
DR	PROSITE; PS00022; EGF_1; UNKNOWN_8.		
DR	PROSITE; PS01186; EGF_2; 7.		
KW	EGF-like domain; Glycoprotein.		
SQ	SEQUENCE 2715 AA; 303062 MW; 598F46A77334C2E1 CRC64;		
Query Match	70.7%;	Score 10309;	DB 11; Length 2715;
Best Local Similarity	67.4%;	Pred. No. 0;	
Matches 1883;	Conservative 378;	Mismatches 395;	Indels 136; Gaps 16;
QY	1	MDVKRR-HRSLTRGCGKECRYTSSLSDEDCRVPTQKSYSSSETLKAYDHD-SRMHYG	58
Db	1	MDVKRRPYCSLTSSKREKERRYTNSSADNEECRVPTQKSYSSSETLKAFDHYSLLYG	60
QY	59	NRVTDLIHRESDEFPGQTNFTLAEIGICEPSPHRS-GVCSMDGILHOGYSLSGSDADS	117
Db	61	NRVKDLVHREADYTRQGNFTLRQLGVCESTRRGVAFCAEMGLPHRGYSISAGSDAT	120
QY	118	DTEGGSPEHAIRLMGRIKSSRSGLSRENSALTLTDSNENKSDDENGRPIPTSP	177
Db	121	ENEAIVMSPEHAMRLMGVKGSRSSCLSSRSNSALTLTOTEHENRSDSEQP-----	173
QY	178	SLLPSAQLPSSINPPVPSQMPLLDSNTSHQIMDTNPDEEFPSPNSVLLLRACSGPQA	237
Db	174	-----SNP-----	177
QY	238	GPNNHHSQTLRPLPPPHNHTLSHHSSANSLNRRSLNRRSQIHA-PAPAPNDLATT	296
Db	178	-----GQPTLQ-PLPPSHKQHPAQHPHSITSLNRRSLNRRNQSPAPPAALPAELQ	230
QY	297	ESVQLQDSVWLVNSVPLETRHFLFKTSSGSTPLFSSSSPGYPLTSGTVYTPPRLPRNT	356
Db	231	ESVQLQDSVWLVNSVPLESRHFLFKTGTGTTPLFSTATPGYTMASGVSYSPTRPLRNT	290
QY	357	FSRKAFKLAKPSKYSWKCAALSAIAALLAILLAYFI-----VPWSL-----	400
Db	291	LSRSAPFKKSKYSWRCTALCAVGVSVLLAILLSYFIAMHLFGLNHLQOTENDTFEN	350
QY	401	-----KNSISDSEAEVGRVVTQVPPGVFWSQIHI 432	
Db	351	GKVNSTVPTNTVSPSGDNKGLGFTHENNTIDSGELDIGRRAIQEVPFGIFWSQLFI	410
QY	433	SQPQLKFNISLGKDALFGVYIRRGPLPPSHAQYDFMERLDG-----KEKWSVESPRE	487
Db	411	DQPQLKFNISLQKDALIGVYGRKGLPPSHQYDFVELLDGSRLLAREQNLVESERAC	470
QY	488	SIOTLVONEAVFQYLDVGLVHLAFYNDGKDKEMYSFNTVLDVSQDCRNCHNGECYS	547
Db	471	QARSVSLHEAGFIQYLDVSGIWHLAFYNDGKNPEQSFNTIVIESVVEECRNCHNGEC	530
QY	548	GYCHFPFGLGADCAKACPVLCSNGQYSGKTCOCYSGWKGAECVPMNQCIDPSCGGH	607
Db	531	GTCHEFPGLGDCSRAACPVLCSNGQYSGKRCCLFCGSKGCTECDVPTTCIDPCGGR	590

QY 608 GSCIDGNCVCSAGYKGBCEBEVDCDPTSCSHGVGVNCECLCSFGMGLNCELARVQCPD 667
Db 591 GICIMGSCACNSGVKGBCEADCDPGCSNHGVCIHGECHCNPGMGWGSNCEILTKWAD 650
QY 668 QCSGHGTYLPDPTGLCSDPNMWDGDCSVGVSDCTHGVCHVIGGACRCEGWGTAACDQR 727
Db 651 QCSGHGTYLQESGCTCDPMTGDCSNEICSDVDCSHGVCMGSCRCCEGWGTPACNQR 710
QY 728 VCHPRCTEHCCTCKDKGCECREGNGEHCCTIGROTAGTETGDCDCLNCGRCRTLQNSWQ 787
Db 711 ACHPRCAEHCTCKDKGCECSOGNGEHCCTIAHYLDKIVKEGCPGLCNSGRCRTLDQNGWH 770
QY 788 CVCOTGHRGPGCNVAMETSCADKNDEGDLVDCDPCCLQACQNSLLCRSGRDLPI 847
Db 771 CVCOPGWRGAGCDVAMETLCTDSKDNEDGLIDCMDDPCCLQSSCQNPYCRGLPDPODI 830
QY 848 IQQG-QT-DMPYKVSFFYDRILKLAGKDSHTHIGENPNFNSLSLIRGVQVYVTDGTPLVG 905
831 ISOSLQTPSOQAASFYDRISFLGSDSTHVLPGESPFSKSLASVIRGQVLTADGTPLLG 890
QY 906 VNVSEFYKPYGYTITRQDGTDFOLIANGASLTALHFERAPFMSOERTVWLPWNSFYAMDT 965
Db 891 VNVSFLHYSEGYTITRQDGMEDLVANGASLTALVFERSFPLQYHTVWIPWNVVYVMDT 950
QY 966 LVNKTENSTPCDLSGFVRPDPITIISSPLSTFTSAAPOGNVPIPEQVLUHEIEELPGSN 1025
Db 951 LVNKKENDIPSCDLSGFVRPSPITVSSPLSTFTSRSSPEDSPIPETQVLUHEIETIPGTD 1010
QY 1026 VKLYLSRSRTAGYKSLKITMTOSTVPLNLRVHLMAVAGHLPKSFQSPASPLASTFIW 1085
Db 1011 LKLSYLSRAAGYKSVUKITMTQAVIPFNLMKVHLMAVAVGRUFQKFPASPPLAYTIFW 1070
QY 1086 DKTDAYQORYGLSDAVSVGYEYETCPSLILWEKRTALLQGFELDPNSLGGWSLDRKHI 1145
Db 1071 DKTDAYNQYVGLSEAVSVGYEYESCLDLTWEKRTAVLQGYELDASNGGWTLDKHHV 1130
QY 1146 LVNKSGLHKGTEGNOPLTOOPAIITSMINGRRRSITCPSGNCGLAEGNKLLAPVALVG 1205
Db 1131 LDVONGTLYKXGNGENOPISQPPVYSSINGRRRSITCPSGNCQADGNGKLLAPVALACG 1190
QY 1206 IDGSLYVGDENYIRIPSPRNVTSLILRNKEFKHNSNPAHYKYLAVDPVSGSLYVSDTN 1265
Db 1191 IDGSLYVGDENYVNRIPSPGNVSVLEARNKDFHSSNPAHYLYLATDPVTDGLYVSDTN 1250
QY 1266 SRRYRYSLSGTDKLAGNESEVAGTGEQCLPFDEARCGDGGKAIDATLMSPRGIAVDKN 1325
1251 TRRYRPKSLTGAKDLTKNAEVVAGTGEQCLPFDEARCGDGGRAVEATLMSPKGMAIDKN 1310
1326 GLMYFVDATMIRKVDQNGIISTLLGSDNLTAVRPLSCDSSMDVAQVREWPTDLAVNPM 1385
Db 1311 GLIYFVDGTMIRKVDQNGIISTLLGSDNLTASARPLTCDTSMHISQVREWPTDLAINPMD 1370
QY 1386 NSLVLENNVILRTENHOVSIITAGRPMHCQVPGIDYSLSKLATHSALESASATAISHTG 1445
Db 1371 NSIIVLNNVYLQITENRQVRIAGRPMHCQVPGVEYIPVGHAVQITLESATATAVSIYG 1430
QY 1446 VLYITETDEKKINRLRQVTTNGEITCLLAGAASDCCKNDVNCNCSGDDAYATDAIINSP 1505
Db 1431 VLYITETDEKKINRQVTTDGEISLVAGIPSECDCKNDANDCQYQSGDGYAKDLNAP 1490
QY 1506 SSLAVAPDGTIYIADLNIRIRAVSKNKPVLNAPNQYEAASPGQEILYVFNADGHIQYTV 1565
Db 1491 SSLASPDGTYLTIADLNIRIRAVSKNKPVLNAPNQYEAASPGQEILYVFNADGHIQYTV 1550
QY 1566 SLVTGEYLYNFTYSTDNDVTELIDNNGNSLKIIRDSSGMPRHLLMPDQNIITLTVGTNGG 1625
Db 1551 SLVTGDLVLYNFSYSDNDVTAFTVDSNGNTLRIRDPNMPVVRVYSPDNQVILWTIGTNGC 1610
QY 1626 LKVYSTQNLELGLMTYDGNLTGLATKSDGTGWTTFYDHEGRLTNTVTRPTGVVYVLSHRE 1685
Db 1611 LKSWTAGLELVLYTYHGNSGLLATKSDGTGWTTFYDSEGRLTNTVTRPTGVVYVLTUHG 1670
QY 1686 MEKSITIDIENNRDDVTVITNLSSVEASYVYVQDVNRNSYQLCNGNGLTRVNVANGMI 1745

Db 1671 MDKAITVDIESSREEDVITSNLSSTDSFTYVWQDLRNSYQIGYDGLSRIFVASGLDS 1730
QY 1746 SFSEHSEHVLAGTITPTIGRCNISLPHENGNSLEWRLRKEQIKGKVITIFGRKLAVHGRNL 1805
Db 1731 HYOTEPHVLAGTANPTVAKRNMTPGNGQNLVEMRFRKEQAOQGVNVFKRKLVRNGNL 1790
QY 1806 LSTDYDNIRITEKTYIDHHRKFTLRIIYDQVRPFLWLPSSGLAAVNVSYFNGRLAGLQR 1865
Db 1791 LSVDFDTTKTEKTYIDHHRKFLRIAYDTSGHPTLWLPSSKLMANVYVYSTGOIASIOR 1850
QY 1866 GAMSERTDIDKQRIVSIRMEFADKQWWSYLYDKSMWLLLSQSQRYIPEYDSSDRLLAVTM 1925
Db 1851 GTTSEKVDYDQGRIVSRVFEADGKTSTYLYEKSWMWLLLSQSQRYIPEYDMDRLSAITM 1910
QY 1926 PSVARHSMSTHSIGYIRNLYNPESNASVIFDYSDGRILTKTSFLGTGRQVFKYKGLS 1985
Db 1911 PSVARHMTQITRISIGYIRNLYNPESNASIITDYNEEGLLQTAFLGTSRRVLFKYRQT 1970
QY 1986 KLSEIVYDSTAVTFGYDETGTGLKMNVLQSGGFSCTIRYRKIGPLVDKQIYREFSEGMVN 2045
Db 1971 RLSEILYDSTRVSTYDETGTAGLTKVNLQSDGFICTIRYRQIGPLIDRQIFRESDGMVN 2030
QY 2046 ARDITYTHDNSPRIASIKPVISETPLPVDLYRYDEISGKVEHFGKFGVYIYDQNIITTA 2105
Db 2031 ARFDYSY-DNSFRVTSMOGVINETPLPIDLYFDDDISGKVEQKFGVYIYDQNIISTA 2089
QY 2106 VMLTSKHEDTHGRKEVQYEMERSLWMTVQYDSMGRVVKRELKLGYPANTTKTYDYD 2165
Db 2090 VMTYKHFDAHGRKEIYQEIFRSLMTWITQYDNMGRVTKREIKGPPANTTKYAYEYD 2149
QY 2166 GDGLOQVAVVNDRPTWRYSYDLNGLNHLNPGNSVRLMPLRYDLDRITRLGDVQYKIDD 2225
Db 2150 VDGOLOVYVLEKIMTWRYNYDLNGLNHLNPPSSARLTPLRYDLDRITRLGDVQYRLDE 2209
QY 2226 DGYLCQBGSDIPEYNSKGLLTRYNKNASGNSVQYRYDCVGRARRSYKTNLGHLLQYFYSD 2285
Db 2210 DGYLRQGTETFEYSKGLLTRYYSKSGWTVIYRYDGLGRRVSSKTSLOQHLLQFFYADL 2269
QY 2286 HNPTRITHVYNSHNSSETSLYDLOGLHFAWESSGSEYVVASDNTGTPLAVFINGLMI 2345
Db 2270 TYPRITHVYNSHNSSETSLYDLOGLHFAWESSGSEYVVASDNTGTPLAVFINGLMI 2329
QY 2346 KQLOYTAYGEIYDSDNPFQMVIGFHGGLYDPLTKLVHFTQRYDYVLAGRWTSPTYTMWK 2405
Db 2330 KQOYTYAYGEIYDSDNPFQMVIGFHGGLYDPLTKLVHFTQRYDYVLAGRWTSPTYTMWK 2389
QY 2406 NVGKEPAPFNLYPKSNPNLSSELDLKNYVDYKSWLVMFGFQLSNLIPIGPFRKMYFVP 2465
Db 2390 RIGKDPAPFNLYPMFRNNNPASKIHDVKDYITDVNSWLVTFGFHLHNAIPGPPVPKFDLTE 2449
QY 2466 PNYELSESQASENGOLITGVQOTTERHNOAFMAL-EGOVITKKLHASTREKAGHWA 2521
Db 2450 PSVELKVSQWEDVPPFVGGVQOQVAROAKAFSLGKMAEVQVSRKAGA---EGSWLWFA 2506
QY 2522 TPTPIIGKIMFAIKRGTVTGVSIIASIESDRKSVASVNNAYILDKMHYSIEGKDTHYFV 2581
Db 2507 TVKSLICKGVMLAVSQGRVQTVNLNLANEDCIVAAVNNAFYLENLHFTIEGKDTHYFI 2566
QY 2582 KIGSADGOLVLTGTTIGRKVLESQVNTVVSQPTLLVNGTRRRRTNIEFOYSTLLLSRYG 2641
Db 2567 KTTTPESDLGTLRLTSGRKALENGINVTVSQSTTVVNGTRRRFADVEMQEGALAHVRYG 2626
QY 2642 LTPDTELEKARVLDQARQALGTAWAKEQKARDGREGSLWTEGKQQLSTGRVQY 2701
Db 2627 M---TLDEEKARILEQARALARAWAREQOQVRDGEGARLWTEGKQQLSTGRVQY 2683
QY 2702 EGYVVLVPEQYPELADSSNITQFLRQNMGRK 2733
Db 2684 DGYVLSVEQYPELADSSANNIQLRQSEIGKR 2715

RESULT 5

Db	59	RVKDLVHHGADEFSGQPDFSLRDMAFGDPVPPHMGAYRTMGLPHRDYSVSVASDADTE	118
Qy	119	TEGMSPEHAIRLWGR-GIKSRSSGSLSRSENSALTLDTSNENKSDODENGRPIPTSSP	177
Db	119	TGIMSPHAIVLWGRNKTSGRSSCLFSRANSNLTLTDTEHEN--TENGPPHLCSSA-	174
Qy	178	SLLPQAQLPSSNNPPV-----SCQPLDSDNTSHQIMDTPDEEFSNLSLLRACSG---	230
Db	175	-----SSSPVDSPPYPPSHAAHQSGRLJLNGSCAQAGRDESEDEFNSFLVKTGSCNVC	230
Qy	231	PQOASSGPPNHHQSOTLR-PLPPPHNHTLSHHH--SSANSLNNSLTLNRRSIOIHAPAPA	288
Db	231	TPAAATANEGSFONHSRLRTPPLPLFHSHSPSQHHTASIGLSHSNNTQRSN-----PSPA	286
Qy	289	PNDLA-----TTPESVOLQDSWLSNVPLETR-----	316
Db	287	PTDSSAPNEGPTSQAODSSAQDNWLLNSVPLETRNIAKOTFLETLODNIEMDILATAR	346
Qy	317	-----HPLFKTSSGSPPLFSSSSPGYPLTSGTVYTPPRLPLPRNTEFSKAPFLKKPS	368
Db	347	RDGAYTDGHLEFK-PGGTSPLYCTTSPGYPLTSTTVYSPPPRPLPRNTEFSRPAFLSKPKY	405
Qy	369	KYCSWKCAALSATAAALLAILLAYFI-----VPWSL-----	400
Db	406	KHCNWKCAALSALLISVTLFFLAYFIAMHFLGNHNLQPVQROIYQITEDNTSGLHLPT	455
Qy	401	-----KNSSIDSGEAEVGRRRVTQEVPPGFWMRSQI	430
Db	466	DLGLPLLGNTGLEFPDRGSRDQKLDGFPFEDSFIDMGEIDVGRKVAQLTPPGFIWRSQV	525
Qy	431	HIISOQFLKFNLSLGDALFGVYIRGLPPSHAQYDFMERLDGKEKWSV-----VESP	483
Db	526	FIDHPMYLAKFNVLSKDALVGYTGRRGALPPSHQTQDFVELLDGRLLLSQGLPLGDGPFPF	585
Qy	484	REBSIOTLVQNEAVFOYLDVGLMHLAFYNDGDKEMVSFNTVVLDSVQDCPNCHNG	543
Db	586	AQORSLVPITSHDTGCIQYMDSGIHWLAIVYDGRKTEQVSFLTATAIDSDDCPSNCFNG	645
Qy	544	ECVSGVCHCFPGPLGADCAKACPVLCSNGQYSGKTCCQYSGWKGAECDPVPMNOCTDPS	603
Db	646	DCVSGNCHCFPGPRGPDCSRACPVLCSSNGQYLAGKCMCHSGWKSGECDVPTNQCIDIT	705
Qy	604	CGHGSGCIDGNCVCSAGYKGEHCEVDCLDPTCSHSGVCVNGECLCSPGWGGGLNCELARV	663
Db	706	CSHGTCIVGTCICNPYSKXGENCEEVDCLDPTCSGSGVCVRGECHCFVGGPGCESPRA	765
Qy	664	QCPDQCSGHCTYLPDFTGLCSDPNWMGPDCSVSVSCVDCGTHGVCIGGACRCEBGWTA	723
Db	766	SCWQCSGHGSGFLADNTNCNHNMTGHDSCSTELCAADCGGHHGICVAGSCRCDGSMWGTG	825
Qy	724	CDQRVCHPRCIEHGTCKDGKCEBCEWNGEHCTIGRQAGTETDGCPLDCLNGNRCRTLGQ	783
Db	826	CEQRACHPRCSEHGTCKDGKCECPWNGEHCIT-----EGCPCLNGNRCRTLGN	876
Qy	784	NSWQCVQOTGWRPGGCVNAVETSCADKNDBEGDLVCLDPPCCLQSAQONSLLRCGRSD	843
Db	877	NGWYVCQLGWRGAGCDTSMETACSDGKDNDDGLTDCMDPOCCLQASCHTTSLCVGS	936
Qy	844	PLDIIOQGO--TDWPAVKSFYDRIKLLACKDSTHIIPGENPFNLSVSLRQGVVTTDGT	901
Db	937	PLDIIOETQISSLSLTSQSFYQRIHFLVGRDSTHVIPDNVDFGTHACVIRGQVYVSDGT	996
Qy	902	PLGVNVSFVKYKGYTTITRODGTFDLIANGASLTLHFERAPMSQERTVWLPWNSFY	961
Db	997	PLGVNVSIFNKPAYGYTTIROGGSFDLVNSGCVAGILURFERAPFTQETHLWLPWGRFF	1056
Qy	962	AMDTLVNKTENSIPSCDLSGFVRPDPIITISSPLSTFFSAAFPQNPVVPETQVLRHEIEL	1021
Db	1057	VMDTIVMRHEVNDIPSCDLSSTFRMPVILPAPLTAFAAGTCPERGIVVPEIQTLOQEVRI	1116
Qy	1022	PGSNVRLYLSSTAGYKSLKLTITWTQSTVPLNLTIRVHLMAVGEHFLQKSFQSPNLAS	1081

Db	1117	POTDMRLGLVSSRSTSGYKSLLRITLTHSTIPPSLMKVKHLMVAVEGRFLFRKWFSAAPNL	SY	1170
Qy	1082	TEIPWKTDAYGORVYGLSDAVYVSVGFEYETCPSLILWEKRTALLAQGFELDPNGLGWSLD	1141	
Db	1177	DFVMDKTDVYSQVYGLSEAFYVSVGFEYBSCPDLLILWEKRTAVLQGYETTASNLGWSVD	1236	
Qy	1142	KHHILNVKSGILHKGTEGNOFLTOQPAITITSMGNGRRRSISCPSCNGLAEGNKLAPVA	1201	
Db	1237	KHHALNIQSGILHKGNGENIFTSQPPVPGTSMGNGRRRSISCPSCNGLADGNKLLAPVA	1296	
Qy	1202	LAVGTDGSLYVGDFNKIRRIFFSRNVTSTLELRNKEFKHNNPAHKYKYLAVDPVSGSLV	1261	
Db	1297	LACGSDGSLYVGDFNKVRRIFTGNVTSLVE-----SNPAHKYKYLATSPVSGWLLV	1349	
Qy	1262	SDTNSRRIRYRKSLGTSKTDLAGNSEYVAGTGQCPLFDEARGCDGGKAKIDATLMSPRGIA	1321	
Db	1350	SDTSSRKVPFKSLYAVDKVAKNLELVAGTGQCPLPYDETRCGDGGKAVEATLTNPRGIT	1409	
Qy	1322	VDKNGLMFPVDATMKRKVDQNGIISTLLGSNDLTAVRPLSCDSSMDVAQVRLEWPTDLAV	1381	
Db	1410	VDKYGVIFEVGDTMRRIDQNGIISTLLGFNDLTSARPLSCDSDVMDISQVRLEWPTDLAV	1469	
Qy	1382	NPMDNSLYLVENNVLIRITENHOVSIAGCPMHQVPGIDYSL-SKLAITHSALESASATA	1440	
Db	1470	SPMDSLYLVDDNNVVLQISENHQVRIVACRPMHQVPGUDHFLVSKIAHTATHLESANALA	1529	
Qy	1441	ISHTGYLITETDEKKINRLQVTTNGEITCLLAGAASDCDKNDVNCNCSGDDAYATDA	1500	
Db	1530	VSHNGLLYIAESDEKKINRVQVSTNGEISLLAGAPSGCDKNDANCDCYSGDDGYAKDA	1589	
Qy	1501	ILNSPSSLAVAPDGTIYIADLGNIRIRAVSKKPVLNAFNQVEAASPGBELYFVNADGI	1560	
Db	1590	KLNAPSLLAVSPDGEILFIADLGNIRIRYVRNKAFLNPLMYEISSPIDDELYLFQVNAS	1649	
Qy	1561	HOYTVSLVTGEYLYNFTYSTDNDVTELIDNNGSLKIRRDSSGMPRHLLMPDNOILTLV	1620	
Db	1650	HVFTQSLTTGDYLYNFTYSGEGDLSITDKNNRVSIRSDTGLPLWLWMPGQGTWFTW	1709	
Qy	1621	GTNGGLKVYSTQNLGLMTYDNGTGLLATKSDETGWTTFYDYDHEGRLTNTVTRPTGVVT	1680	
Db	1710	GTNNALKSVAAQGEIATVMTYHGSSGLLATKSNEDGWSTFEYDNYGRLTNTVYTPGRVS	1769	
Qy	1681	SLHREMEKSIITDIENSNRDDVTVTNLSSVEASTVYVQDVQRNSYQICLNCTLRVMTA	1740	
Db	1770	SYRTSDSTVRVQTEGSKN-EDITVTNLSSAGSTFYTLMDQOVKNSYYIGLDGSLRLVA	1828	
Qy	1741	NGMGTSFHSEPHVLACTIPTIGRCNISLMPENGLNASIEWRLRKEOIKGKVTIFGRKLVR	1800	
Db	1829	NGMEVSLHTPEHLUSCTVNPSTSKRNTYIPIDNGLNLVEMQRKEAQGVTVYGRRLVR	1888	
Qy	1801	HGRNLLSIDYDRNIRTEKTYDHRKFTLRIYDQVGRPFLLWPLSSGLAAVNVSYFFNGRL	1860	
Db	1889	HNRNLLSMDFDVRTKEKVVYDHRKFTLRIHYDHAGRPTLWAPSSRLNGVNVYTPSGGHI	1948	
Qy	1861	AGLQCGAMSERTDIDKQGRIVSMPADGKWYSYLDKSMVLLLSQSQRYIEFYDSSDRL	1920	
Db	1949	AGIQRTGMSVRMEYDQNGRITSKIFADGKWSYTYLEKSNWLLLYSQROYIEFEFKNDRL	2008	
Qy	1921	LAVTWPSVARHSMSTHTSTGYTRNIYNPPENASVTFDYSDCGRIILKTSFLATGROVEYK	1980	
Db	2009	SVYTMPNVAARQLETTTRTSIGYRYNTYRPEPNATVLDQYSEDGLLQOTIQHGGRRVYK	2068	
Qy	1981	YGLKSLSEIVYDSTAVTGYDEYTTGVLKMNVLQSGGFCTTRYRKIGPLVDKQIYRFSE	2040	
Db	2069	YGLSLRLEILLYDTTRIAFSYDESAGMLTVGLQSGGFACITRYRQIGPLIDRQIFRFE	2128	
Qy	2041	EGMVNARFDYTHDNSFIASIKPVISETPLPVDLYRYDEISGKVEHFGKFGVYYDINO	2100	
Db	2129	EGMVNARFDYK-DNSFRVTSMQAVINETPLPIDLYRYDDVSGKTEQFGKFGVYYDINO	2187	
Qy	2101	IITTAVMTLKSLHFDTHGRKEIYQVEMFRSLMYMTVQYDSMGRIYIRKELKGPYANTYKY	2160	
Db	2188	IITTAVMTHTKHFDAYGRYKEVQYETFRSLMYMMVQFNMGRVAKELKVGYPYANTY	2247	

QY	2161	TYDYGDGQLQSVAVNDRPTWRYSYDLNGLNHLNPFNSVRLMPLRYLDRITRLGDVQ	2220
Db	2248	AYEYDADGQLQVYSINDKPLWRYSYDLNGLNHLSPGNSARLTPLRYDIRITRLGDVQ	2307
QY	2221	YKIDDDGYLCQSGSDIFENYKGLLTRYANKASGWSQVRYDGVGRASYKTNLGHLLQY	2280
Db	2308	YRLDEGFLQRGNDFEFENYKGLLTRYANKASGWSQVRYDGVGRASYKTNLGHLLQY	2367
QY	2281	FYSDLNHPTRITHVYHNHNSSEITSLYDLOGHLFAMESSGREYVVASDNTGPTPLAVESI	2340
Db	2368	FYADLSPTRITHVYHNHNSSEITSLYDLOGHLFAMESSGREYVVASDNTGPTPLAVESI	2427
QY	2341	NGLMIKQLQYATGEIYYDSNPDMQVIGFHGGLYDPLTKLVHFTQRDXYDLVLAGRWTS	2400
Db	2428	AGLMIKQLHTAFTGEIYYDSNPDMQVIGFHGGLYDPLTKLVHFTQRDXYDLVLAGRWTS	2487
QY	2401	YTMKWKNGKEP-APPNLYMFKSNPNLSSELDLKNYVTDVYKSWLVMFGFQLSNIIPGPRA	2459
Db	2488	HDIRKRLNSDNIVPFNLYMFKSNPNLSNQETKCYMTDVSMLVTFGFQLYNVIPGRKP	2547
QY	2460	KMYFVPPVPELSQAS-ENGOLITGVQOQTERHNOAFMALE--GOVITKK-----	2507
Db	2548	VTDAMPSTVELVHTQKTQEWDSKSVLGVQCEVQRLKSFVRLERFGQIYASDSGCPP	2607
QY	2508	--LHASIREKAGHWFATTPIIGKGMFAIKRGRVTTGVSSIASSEDSRKYVASVNNAYYL	2565
Db	2608	TPLHT-----LFGTGLSFGKGVKVAIREGRVADIIISLANEDGRRIAAVLDKASYL	2659
QY	2566	DKMHSIEGKDHIFYFKISADGDIYTLGTTGRKVLKESGVNVTVSQPTLLVNGRTRRT	2625
Db	2660	QDLHFTIAGLDTHIFYKSLVEGDLTLGNTVQGTLETGVNVTVSQVNVLGGRRRT	2719
QY	2626	NIEFOYSTLLSTRYGLTPTDILDEEKARYLDOARALGTAWAKEOQKARDGREGSLWT	2685
Db	2720	DIOMQVGTLSLVRYG---SSVDEEKRVVLELQARAVATATAWAHERHRLRQEGEGRWT	2776
QY	2686	EKEQQLLSTGRVQGYEGYVLPVEQYPELADSSNIQFLRQNEGKR	2733
Db	2777	DGERQQLSSGRVQGYEGYVIVSDQFPELTINNHNHWEQTEMGR	2824
RESULT	9		
Q9JLC1			
ID	Q9JLC1	PRELIMINARY;	PRT; 2346 AA.
AC	Q9JLC1;		
DT	01-OCT-2000 (Tremblrel. 15, Created)		
DT	01-OCT-2000 (Tremblrel. 15, Last sequence update)		
DT	01-Dec-2001 (Tremblrel. 19, Last annotation update)		
OD23	(FRAGMENT).		
OD23.			
Mus	musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxId=10090;		
RN	[1]		
RA	SEQUENCE FROM N.A.		
RP	Ben-zur T., Felge E., Motro B., Wides R.;		
RT	"The mammalian Odz gene family: Homologs of a Drosophila pair rule		
RT	gene with expression implying distinct yet overlapping developmental		
RT	roles.";		
RL	Dev. Biol. 117:107-120(2000).		
DR	EMBL; AF195416; AAF28316.1; -.		
DR	HSSP; P01492; LVNB.		
DR	InterPro; IPR000561; EGF-like.		
DR	InterPro; IPR002049; Laminin_EGF.		
DR	Pfam; PF00008; EGF; 6.		
DR	PRINTS; PR00011; EGF_LAMININ.		
DR	SMART; SM00181; EGF; 5.		
DR	PROSITE; PS00022; EGF_1; UNKNOWN_8.		
DR	PROSITE; PS01186; EGF_2; 7.		
KW	EGF-like domain; Glycoprotein.		
FT	NON_TER	1	1

SQ	SEQUENCE	2346 AA;	261699 MW;	D761563C6AEDD735 CRC64;
	Query Match	63.2%;	Score 9209;	DB 11; Length 2346;
	Best Local Similarity	70.7%;	Pred. No. 0;	
	Matches 1657;	Conservative 327;	Mismatches 326;	Indels 34; Gaps 9;
QY	401	KNSSIDGEAEVGRRTVQVPPGVFWRSQIHSQPOFLAKFNISLQKDALFGVYIRGLPP	460	
Db	26	ENNTIDSGELDIGRAIQVPPGVFWRSQLFDPQFLAKFNISLQKDALIGVYGRGLPP	85	
QY	461	SHAQYDFMERLDG----KEKNSVSPRRRSIQTLVQNEAVFYQYLDVGLMHLAFYND	515	
Db	86	SHTQYDEVLLDGSRLTAREQRLVSESRAGROARSVLSHEAGFIQYLDGSIWHLAFYND	145	
QY	516	GKDKEMSVNTVVLDSVODCPRNCHNGECVSCVCHCFGLGADCAKACAPVLCSGNGQ	575	
Db	146	GKPEQVSENTIVIESVVECPRNCHNGECVSGTCHCFGLGADCAKACAPVLCSGNGQ	205	
QY	576	YSKGTCCQYSGWKAGCDVPMNQCIDPSGGHSGCIDGNCVCAGYKGBHCEVDCLDPT	635	
Db	206	YSKGRCLCFSGWKGTEDVPTTQCIDPQCGRGICIMGSCACNSGYKGENCEADCLDPC	265	
QY	636	CSSHGVCVNGECLSGPWGGLNCELARVQCPDQCSGHGYLPTGTLCSDPNMNGPDCSV	695	
Db	266	CNSHGVCIHGECHNCPGWGSGNCEILKTMCADQCSGHGYLQESGCTCDPNMTGDCSN	325	
QY	696	EYCSVDCGTHGVCIGACRCEEGTGAACDORVCHPCIEHGTCCKGCKCECEGNGEHC	755	
Db	326	EYCSVDCGSHGVCMGSGRCEEGTGPACNORACHPCAEHGTCCKGCKCECEGNGEHC	385	
QY	756	TIGROTAGTETDGPCLDNGNRCITLQNSWCVCQGTGWRGPGCNVAMETSCADNKDNEG	815	
Db	386	TI-----ECCPGLCNSNGRCTLQNGHWCVCQPGWRGACGCDVAMETLCTDSKDNEG	436	
QY	816	DGLVCLDLPDCLQSAQNSLLCRGSRDPLDIIQQG-QT-DWPAVKSFFDRIKLLAGKDS	873	
Db	437	DGLIDCMDPCLQSSCQNPYCRGLPDQDIISQSLQTPSQQAASFYDRISFLSGDS	496	
QY	874	THIIPENPNSSLIRGVVTTDGTPLVGVNVFVKYPKYGYTTQDGTGFDLIANG	933	
Db	497	THVLPGESPFNSLASVIRGOVLTDGTPPLGVNVFVFLHYSEYGYTTQDGTGFDLVANG	556	
QY	934	GASLTUHFERAPMSQERTVWLPWNSFYAMDTLVMKTEENSPSCDLSGFSVRPDPPIISS	993	
Db	557	GASLTUHFERSPELTQYHTWIPWNVFYMDTLVMKKEENDIPSCDLSGFSVRPSPIVSS	616	
QY	994	PLSTFFSAAPGQNPVTPQVHLHEELPQSNVVKLRYLSSRTAGYKSLKITMTQSTVPL	1053	
Db	617	PLSTFFSRSPEDSPIIPTQVHLHEETTPGTDLKLSYLSRAAGYSVKLKITMTQAVIPF	676	
QY	1054	NLIRVHLMVAVEGHLFQKSFQASPNLASFIMDKTDAYGORYVGLSDAVYVSGFEYETCP	1113	
Db	677	NLMKVHLMVAVYGRLEFQKWFPPASPFLAYTFIMDKTDAYNQYVGLSEAVYVSGFEYESCL	736	
QY	1114	SLILWEKRTALQGFELDPNSLGGSLDKHHILNVKSGILHKGTGENQFELTQPAITSI	1173	
Db	737	DLTLWEKRTAVLQGYELDASNMGGWTLDKHHVLDVQNGILYKNGENQFISQPPVSSI	796	
QY	1174	MNGRRRSISCSNGLAEGNKLLAPVALAVGIDGSLYVCDENYIRIIPSRNVTSILEL	1233	
Db	797	MNGRRRSISCSNCGAQDGNKLLAPVALACGIDGSLYVCDENYIRIIPSRNVTSILEL	856	
QY	1234	RNKEFKHSNNPAHYLYLAVDPVSGSLYVSDTNSRRIRYVRKSLSGTKDLAGNSVWAGTGE	1293	
Db	857	-----SSNPAHYLYLATDPTVTDLYVSDTNTRRIRYVRKSLSGTKDLKNAEVWAGTGE	909	
QY	1294	QCLPFDEARCGGKKAIDATLMSPRGIAVDKNGIMTFVDMTRKVDQNGIISTLGSND	1353	
Db	910	QCLPFDEARCGGKRAVEATLMSPKGMAIDKNGIYFVDMTRKVDQNGIISTLGSND	969	
QY	1354	LTAVRPLSCDSSMDVAQVLEWPTDLAVPMDSLVLENVILRLITENHOVSIAGRPM	1413	

Db 970 LTSARPLTCDTSMHISQVRLEWPTDLAINPMDNSIYYLDNNVVLQITENROVRIAAGRPM 1029
Qy 1414 HCQVPGIDYSLKLAHSALESASAIASHTGVLVYITETDEKKNRLRQVTTNGEICLLA 1473
Db 1030 HCQVPGVEYVPGKHAVQTTLESATAIAVSYSGVLVYITETDEKKNRLRQVTTNGEISIVA 1089
Qy 1474 GAASDCCKNDVNCNYSGGDAYATDAILNSPSSSLAVAPDGTIYIADLGNIRIRAVSKNK 1533
Db 1090 GIPSECDCKNDANDCYQSGDGYAKDARLAPSSSLAASPDGTLVYIADLGNIRIRAVSKNK 1149
Qy 1534 PVLNAFNOYEASPGCEGLYVFNADGTHOYTVSLVGTGEYLYNFYSTDNDVTELIDNNGN 1593
Db 1150 PLLNSMFYVASPTDQELIFIDNGHQYTVSLVGTGDIYFNFSNDNDVATVDSNGN 1209
Qy 1594 SLKTRRSSGMPRHLLPDNDQIITLVGTNGGLKAVSWSTQNLLEGLMTYDNGTGLLATKSD 1653
Db 1210 TLRIRDPNRPVVRVSPDNQVILWLTGTNGCLAKSMTAQGLELVLYFHGNSGLLATKSD 1269
Db 1654 ETGWTTTFDYDHEGRLNVRPTGVVTSLSHREMEKSITIDIENSNRDDDDVTVITNLSVSE 1713
Db 1270 ETGWTTTFDYDSEGLTNVPTGVTNHLGDMDKAITVDIESSSREEDVSITSNLSIID 1329
Qy 1714 ASYTVVODVNSYQLCNNGCTFLRVYANGMGISPHSEPHVLAGTITPTIGRCNLSLPMEN 1773
Db 1330 SFYTWQDLRNSYQIGVDSGLRIFYASGLDSHTQTEPHVLAGTANPTVAKRNMTPGPN 1389
Qy 1774 GLNSIELRLRKEQIKGVITFGRKLRVHGRNLLSIDYDRNIRTEKIIDDDHRKFTLRIYD 1833
Db 1390 GONLVENFRERKEQAQGVNVFGRKLRVNGRNLSSVDFDRTTKTEKIIDDDHRKFTLRIYD 1449
Qy 1834 QVGRPFWLSPSSGLAANVSYFFNGRLAGLQAGMSEKRTDQGRIVSRMFADGKVMYS 1893
Db 1450 TSGHPTLWLPSSKLMANVTYSSTGQIASQRTGTEKVDYDSQGRIVSRMFADGKVMYS 1509
Qy 1894 SYLBSKMWLLQSORQYIFEDYSDRLAVTMPVSVARHSMSTHTSIGYRNIYNPPESNA 1953
Db 1510 TYLEKSMVLLHSORQYIFEDYDMDRLSAITMPSVARHTWOTIRSIGYRNIYNPPESNA 1569
Qy 1954 SVIFDYSDGGRILKTSFLGTGRQVYFYKGLSKLSEIYDSTAVTFGYDETGTGLKVMNL 2013
Db 1570 SIITDYNEEGLLQATFLGTSSRLVFLYRQTRTLEIYDSTRVSFTYDAGVLTKNL 1629
Qy 2014 QSGFSCITIRYRKGLPLVDKQIYRFSEGGWYNARFDYTHDNFSRIASIKPVISETPLPV 2073
Db 1630 QSDGFICTIRYRQGLPLDRIQIFRESDGMVYARFDSY-DNSFRVTSMQGVINETPLPI 1688
Qy 2074 DLYRYDEISGRVHFGRGVYIYDINOITTAVMTLKSKHFDTHGRKEVQYEMFERSLMYW 2133
Db 1689 DLYQFDDISGRVHFGRGVYIYDINOITTAVMTLKSKHFDTHGRKEVQYEMFERSLMYW 1748
Qy 2134 MTVOYDSMGRVIRKRELKGPYANTTKYTYDYDGDGQLQSVAVNDORPTWRYSYDLNGLNHL 2193
Db 1749 ITIOYDNMGRTYKREIKIGPANTTKYAYEYDVGQLQTVYLNKIMWRYNDLNGNLHL 1808
Qy 2194 LNPNSVRLMPLRYDLDRITRLGDVQYKIDDDGYLCQSGDIEFYNSKGLLTRAYNKAS 2253
Db 1809 LNPSSARLTPLRYDLDRITRLGDVQYRLDEGDFLQRGTEIFEYSKGLLTRAYSKGS 1868
Qy 2254 GWSVOYRYDVGGRASVYKTLGHHLQVYSDLHNPTRITHVYNSNSNITSLSLYDLQGLH 2313
Db 1869 GWTVIYRYDGLGRRVSSKTSUGHLQFFYADLTPTPTITHVYNSNSNITSLSLYDLQGLH 1928
Qy 2314 FAMESSSGEYVVASDNTGTPLAVFSINGLMIKQLQYATYGEIYDSDPDMFOMVIGFHG 2373
Db 1929 FAMELSSGDEYIASDNTGTPLAVFSSNGLMKQIYATYGEIYDSDNVDFOLVIGFHG 1988
Qy 2374 LYDPLTKLVHFQTDYDVLAGRWTSPDYTMWKNVKGEPAPNLNMYFNKSNPNPLSSSELDLKN 2433
Db 1989 LYDPLTKLVHFGERDYDILAGRWTTPDTEIKRIGKDPAPENLYMFRNNPNASKIHDVKD 2048
Qy 2434 YTVDWKSHLVNMFQGLSNIIFGFPRAKMYFVPPPYELSESQASQNLITGVQQTTERHN 2493
Db 2049 YITDYNLSLVTFGFHLHNAIFGFPVPKFDLTPEPSYELVKSQOWEDVPPIFGVQOQVARGA 2108

Qy 2494 QAFMAL----EGQVITRKLHASIREKAGHWFPATTTPIIGKIMFAIREGRVTTGVSSIAS 2549
Db 2109 KAFSLGCKMAEVQVSRKAGA--EQSWLWFATVKSLLIGKVMGLAVSGRQVOTVNLNAN 2165
Qy 2550 EDSKVASVLNAYYLDKMHYSIEGKOTHYFVKTGSADGDLVTLGTTIGRKRVLSGVNVT 2609
Db 2166 EDCIKVAALNNAPYLENLHFTIEGKOTHYFVKTTPTESDLGTLRLTSGRKALENGINVT 2225
Qy 2610 VSOPTLLVNGTRFRFTNIEFOYSTLLLSIRYGLTPDTLDEEKARVLDQARORALGTAWK 2669
Db 2226 VSQSTTVVNGTRFRFADVEMQFGALAHVRVGM--TLDEEKARLEQARORALARAWAR 2282
Qy 2670 EQQRVDRGEGARLWTEGEKRLQQLSGRQVQEGYGVYVLPVQYPELADSSNIQFLRQNE 2729
Db 2283 EQQRVDRGEGARLWTEGEKRLQQLSGRQVQEGYGVYVLPVQYPELADSSNIQFLRQNE 2342
Qy 2730 MKGR 2733
Db 2343 IGKR 2346

RESULT 10
Q9ULU2
ID Q9ULU2 PRELIMINARY; PRT; 1737 AA.
AC Q9ULU2;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE KIAA1127 PROTEIN (FRAGMENT).
GN KIAA1127.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=20039618; PubMed=10574461;
RA Hirosewa M., Nagase T., Ishikawa K., Kikuno R., Nomura N., Ohara O.;
RT "Characterization of cDNA clones selected by the Genemark analysis
from size-fractionated cDNA libraries from human brain.";
RL DNA Res. 6:329-336(1999).
DR EMBL; AB032953; BAA86441.1; -.
DR InterPro; IPR001258; NHL.
DR Pfam; PF01436; NHL; 2.
FT NON_TER 1
SQ SEQUENCE 1737 AA; 195065 MW; CFF0AF0C8C5D6B32 CRC64;

Query Match 62.1%; Score 9055; DB 4; Length 1737;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1736; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 997 TFFSAAPQNPVETQVLEHEEELPGSNVLRVLSSTAGYKSLKTKTMTQSTVPLNLI 1056
Db 1 TFFSAAPQNPVETQVLEHEEELPGSNVLRVLSSTAGYKSLKTKTMTQSTVPLNLI 60
Qy 1057 RVHLMVAVEGHFLQKSFQASPNLASTFTWDKTDAYGQRYGLSDAVSVSGFEYETCPSLI 1116
Db 61 RVHLMVAVEGHFLQKSFQASPNLAYFTWDKTDAYGQRYGLSDAVSVSGFEYETCPSLI 120
Qy 1117 LWKRTALLQCFELDPNSNLGWSLDKHHILNVKSGILHKGTTGENOFTQQPAITISIMGN 1176
Db 121 LWKRTALLQCFELDPNSNLGWSLDKHHILNVKSGILHKGTTGENOFTQQPAITISIMGN 180
Qy 1177 GRRSISPCSCNGLAEGKNLAPVALAVGIDGSLYVGDFNYIRRIFFPSRNVTSILELRNK 1236
Db 181 GRRSISPCSCNGLAEGKNLAPVALAVGIDGSLYVGDFNYIRRIFFPSRNVTSILELRNK 240
Qy 1237 EFKHSNNPAHKYIYLAADVPSGSLYVSDTNSRRIYRVKSLGKDLAGNSEVAVAGTGEOL 1296
Db 241 EFKHSNNPAHKYIYLAADVPSGSLYVSDTNSRRIYRVKSLGKDLAGNSEVAVAGTGEOL 300

Db 904 NVSFLHSDYGTISRQDGSFDLVAIGGISVVLIFDRSPFLSEKRTLWLPNQFIIVEKV 963
Qy 967 VMKTEENSIPSCDLSGFRVPPPIIISPLSTFESAAPQNPVPEPTOVLEHEIEELPGSNV 1026
Db 964 IMQIRIVADAPSCDLSNFIISPNIPLVPSLTFGGSCPRGTHIVFELQVYQVEIPIPSFV 1023
Qy 1027 KLRLYSSRTAGYKSLTKITMTOSTVPLNLIRLHVLMVAVFEGHLFOFSQASPNLSTFIWD 1086
Db 1024 RLSYLSSRTPCYKTLRLTLRLLTHSTIPVGMKIVHLTVSVEGLRTQKWFPAALNLVYTFAWN 1083
Qy 1087 KTDAYGQRYVGLSNAVSVSGEYETPCSLIILWEKRTALLOFELDPSPNLGWSLDDKHIL 1146
Db 1084 KTDYIGQKRWGLAEALVSGVEYEMCFILWEQRTVYVLOGFEMDASNLGWSLKNKHIF 1143
Qy 1147 NVKSGILHKGTEGNOFTQOPAITTSIMNGRRRSISCPSCNGLAEGKNLLAPVALAVGI 1206
Db 1144 NPGSIIHKNGENWFISQPPVATATMGCHQHSVACTNCGPAHNNKLPAPVALASGP 1203
Qy 1207 DGSLYVGFNFIYRIRFPSSNVTSLERLNEKFNHNNPAHKYILAVDPVSGSLXVSOTNS 1266
Db 1204 DGSVVGDFNFIYRIRFPSSNVTSLERLNRDTRHSTSPAHKYILAMDPMSESLYLSOTNT 1263
Qy 1267 RRIYRVKSLCTKDLAGNSEVYAGTGEQCLPFDEARCGDGKKAIDATLMSRGLTAVDKNG 1326
Db 1264 RKVYKLSLVETKDLKSNFEVYAGTGDQCLPFDSCHGDGKSEASLNSPRGITVDRHG 1323
Qy 1327 LMIFYDATMIRKVDQNGIISTLLGSLNDLTAVRPLSCDSMDVAOVRLWEPMTDLAVNPMDN 1386
Db 1324 FIYFVGTMIARIDENAVITVIGSGLTSTQPLSCDSGMDITQVRLWEPMTDLAVNPMDN 1383
Qy 1387 SLYVLENNVILRITENHOVSIIAGRPMHCQVPGIDYSL- SKLATHSALESASAIASHTG 1445
Db 1384 SLYVLDNNVILQISENNRVRRIAGRPICQVPGIDHFLVSKVAITHSTLESARAISSVSHG 1443
Qy 1446 VLYTETDEKKINLRQVTTNGEICLLAGASDCCKNDVNCNCSGDDATATDAILNSP 1505
Db 1444 LLFIAETDERKVNRIQVTTNGEISIIAGAPTDCCKIDPNCDFSGDGGYAKAKMKAP 1503
Qy 1506 SSLAVAPDGTIADLGNIRAVSKNPVLNFAFNQYEAASPGEOELVFNADGTHQVTV 1565
Db 1504 SSLAVSPDGLVADLGNVIRITISKNOHLNDMNLIEIASPADEQLYQFTVNGTHLTM 1563
Qy 1566 SLVGEYLYNTYSTDNVDTELIDNNGSLKIRDDSSCPHRLMDPQNIITLTVGTNGG 1625
Db 1564 NLITRDYVYNYETNAEGDLGATSSNGSNVHIRRAGGMPLWLVPVGGQVYWLATSSNGV 1623
Qy 1626 LKVYSTONLEGLMTYDGTGLLTKSDGTGTTTFYDYDHEGRLTNVTRPTGVVTSLSHRE 1685
Db 1624 LKRVSAQYNNALMTYPCNTGLLTKSNENGWTTVYEDPEGHLTNATFPTGEVSSFSHD 1683
Qy 1686 MEKSTIDIENSRDDDDVTITNLSSVEASTYVQDOVRNSYQCLNNGTLRVMYANGMI 1745
Db 1684 LEKLTQKVALDYSNR-ENVLMSTNLATSTIVILAKQENTQSTYRVSPDGLSLRVTFAAGMEI 1742
Qy 1746 SFHSEPHVLVAGTPTTGRCNLSLPMENGLSIEWRLKKEIKGKVTIFGRKLRLVHGRL 1805
Db 1743 NLSSEPHILAGAVNPPLTKGKCNISLPGEHANLIEWRQKEQNGKVSAPAFERRLAHNNL 1802
Qy 1806 LSIDYDRNIRTEKIIDHDKFTLRIYDQVGRPFPLWLPSSGLAAVNVSYFENGLRAGLQR 1865
Db 1803 LSIDFDMTRTKGIYDDHDKFTLRIYDQGRPLWSPVRYNEVNTIYSPGLVTFLLQR 1862
Qy 1866 GAMSERTIDKOGRIVSRMFADGKWSYSYLDKSMVLLQSORQVIFYEYDSSDRLLAVTM 1925
Db 1863 GTWNEKMEYDQSGKIISRTWADGKIWSYLEKSMVLLHLSQRRYIFYEYDQSDCLLSVTM 1922
Qy 1926 PSVARHSMSTHTSGYIRININYPSPSNASVTFDYSDDGRILKTSPLGTGRQVYFYKGLS 1985
Db 1923 PSWVRHSLQTLMSVGYRYNIYTPPDSSTSFIDQYSRDRGLLQTLHLGTGRRVLYKYTKQA 1982
Qy 1986 KLSIETVYSTAVTFCYDETTGVLMKVNLOSQGSFCTIYRKIGPLVDKQIYRFSEGMVN 2045
Db 1983 RLSEILYDTTQVLTLYEESSGVIKTIHLMDHGFICTIRYRTGTGPIGRQIFRFSSEGLVN 2042

Qy 2046 ARFDYTYHDNSFRIASIKPVISETPLPVDLYRYDEISGKVEHFGVGYVYDIHQIITTA 2105
Db 2043 ARFDYSY--NNFRTYSMQAVINETPLPIDLYRYVDVDSRTQFQKFSVINDLQNOVITT 2100
Qy 2106 VMTLSKHFDTHGRKEVOYEMFRSLMYMTVOYDSMGRVIRKREIKLGPYANTKYTYDYD 2165
Db 2101 VMKTKIFNANGQVTEVQYELKAIAYWMTIQYDNMGRVICTDIRGVGDVANTIRYFYEYD 2160
Qy 2166 GDGQLOSVAVANDRTWRYSYDLNGLNHLNPGNSVRLMPLRYDLRDRITRLGDVQYKIDD 2225
Db 2161 ADGQLOTYSVNDKIQWRYSYDLNGLNHLNPGNSVRLMPLRYDLRDRITRLGEIQYKNDE 2220
Qy 2226 DGYLCQSGSDTIFEYNSKGLLTRAYNKASGWSQVRYDCVGRRASVKTNLGHHLQVYFSDL 2285
Db 2221 DGFLRQRGNDLIFEYNSKGLLTRAYNKASGWSQVRYDCVGRRASVKTNLGHHLQVYFSDL 2280
Qy 2286 HNPTRITHVYNSHSETSLYDLOGHLFAMESSGEYVYVASDNTGTPPLAVFSINGLMI 2345
Db 2281 ANPIRVTHLYNHTSAETISLYYDLQGLHILAMELSGEEYVYVACDNMGTPPLAVFSRQVVI 2340
Qy 2346 KOLOYTAYGEIYDSDNPDQMVIGFPHGGLYDPLTKLVHFTORDYDVLAGRWTSPDYTMWK 2405
Db 2341 KEILYTPGDIYHDTPDEFVIIGFPHGGLYDPLTKLVHFTORDYDVLAGRWTTPNHHLWK 2400
Qy 2406 NVGKEPAPFNLYMFKSNPLSSELDKNYVTDVKSWSLVMFGFQSLNSIIPGFPRAKMYFVP 2465
Db 2401 QNLNLPKPNLYSPENNYPVGKIQDVAKYTTDGTWLELFGQLHNLVPGFPKPELENME 2460
Qy 2466 PPYEL-----SSQASENQLITGVQOOTTERHNOAPMALEGOVITKKLHASTREKAGHW-- 2519
Db 2461 LTYELQLQOTKTDWDPGKMILGICQELQKQLRNFISLDQLPMTQYNEGRGLEGGKQPR 2520
Qy 2520 FATTPITGKIGMFAKEGRVTTGVSSTASEDSRKAASVLANNAYVLDKMHVYSIEGKDPHY 2579
Db 2521 FAAPSVFGKIGKFAKEGIVTADIIGVANEDESRKLAAILNNAHLENLHETIEGRDTHY 2580
Qy 2580 FVKIGSADGLVTGTTIGRKVLESGVNVTVSQPTLLVNGRTRFRFTNIEFOYSTLLLSIR 2639
Db 2581 FIKLSLEEDLVLIGTGGRRILENGVNVTVSQMVSVLNGRTRFRFADIQHGLGCFNIR 2640
Qy 2640 YGLPDPDLDEKARVLDQARALGTAWAKEQKARDREGSRRLWTEGEKQOLLSTGRVQ 2699
Db 2641 YG---TTVEEKHNHVENARQAVAQVTEORLQEGEGEGRVWTEGEKQOLLSTGRVQ 2697
Qy 2700 GYEGYVLPVQYPELADSSNIQFLRONEMGR 2733
Db 2698 GYDGYFVLSVEQYLELSDSANNIHPMRQSEIGRR 2731

RESULT 13
Q9UKZ4
ID Q9UKZ4 PRELIMINARY; PRY; 2725 AA.
AC Q9UKZ4;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE TENASCIN-M1.
GN TNM1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N. A.
RX MEDLINE-20025749; PubMed-10556288;
RA Brandau O., Schuster V., Weiss M., Hellebrand H., Fink F.M.,
RA Kreczy A., Friedrich W., Strahm B., Niemeyer C., Belohradsky B.H.,
RA Meindl A.;
RT "Epstein-Barr virus-negative boys with non-Hodgkin lymphoma are
RT mutated in the SH2D1A gene, as are patients with X-linked
RT lymphoproliferative disease (XLP).";
RL Hum. Mol. Genet. 8:2407-2413(1999).

Qy 1820 YDDRHKFTLRIYDQVGRPELWSPSSGLAAVNVSYFFNGRLAGLQRGAMSBRTDIDKQGR 1879
Db 1811 YDDRHKFTLRIYDQVGRPELWSPSSGLAAVNVSYFFNGRLAGLQRGAMSBRTDIDKQGR 1870
Qy 1880 IVSRMFADGKWSYSLDKSWLLQSQORQYIFEDSDSDRLAVLTPMSPVAHSMSTHTSI 1939
Db 1871 IISRTWADGKWSYSLDKSWLLQSQORQYIFEDSDSDRLAVLTPMSPVAHSMSTHTSI 1930
Qy 1940 GYIRNIYNPESNASVYDFVSDGRIKLTFLGTGROVEYKYGKSLSEIVDYDSTAVTF 1999
Db 1931 GYIRNIYNPESNASVYDFVSDGRIKLTFLGTGROVEYKYGKSLSEIVDYDSTAVTF 1990
Qy 2000 GYDETTGVLKMNLOSQGFCTIIRYKIGPLVDQKQIYRFSEEGMVNAREFYTHDNFSRI 2059
Db 1991 TYEESGVIKTHLMDGFCITIRYQGTGLIGRIFRSEEGVLNARFDSY--NNFRV 2048
Qy 2060 ASIKPVISETPLVDLYRYDEISKVEHFGKFGVYIYDINOIITAVMTLSKHEDTHGRI 2119
Db 2049 TSMQAVINETPLVDLYRYDVSQRTQFGKFSVINDLQNOVITTTVKKHTKIFSANGQV 2108
Qy 2120 KEVOYEMFRSLMYWMTVOYDSMGRIKRELKPLGYANTTKYTYDYDGDGQLQSVAVANDRP 2179
Db 2109 IEVOYEILKATAYWMTVOYDNVGRHNCIRGVYDANITRYEYDADGQLQTVSYNDKT 2168
Qy 2180 TWYSDYDLNGLHLNPGNSVRLMPLRYDLRDRITRLGDVOYKIDDDGYLCQSGSDIFEY 2239
Db 2169 QWYSDYDLNGLHLNPGNSVRLMPLRYDLRDRITRLGEIQYKMDGDFLRQGNDFEY 2228
Qy 2240 NSKGLLFRAYNKASGWSVOYRYDGVGRASYKTNLGHHLQYFSDLNHPTRITHVYHNSN 2299
Db 2229 NSNGLLQKAYNKASGWSVOYRYDGVGRASYKTNLGHHLQYFSDLNHPTRITHVYHNSN 2288
Qy 2300 SEITSLYDQGLHIFAMESSEGEYVYASDNTGTPLAVFSINGLMIKOLOYATYGEIYD 2359
Db 2289 SEITSLYDQGLHIFAMESSEGEYVYACDNTGTPLAVFSINGLMIKOLOYATYGEIYD 2348
Qy 2360 SNPDFQVIGPHGLYDPLTKLVHFTQDYDLVLAGRMTSPDYTWKNNVKGEPAPFNLYMF 2419
Db 2349 TYPDFQVIGPHGLYDPLTKLVHFTQDYDLVLAGRMTSPDYTWKNNVKGEPAPFNLYMF 2408
Qy 2420 KSNPLSELKLVNVDVKSVMVFGFOLSNIIPGPRAKMYFVPPPYEL---SESQA 2475
Db 2409 ENNPVKGIDQVAKYTTDIRSWLELFGQLHNVLPGPKPELENLELYELLRLQTKQE 2468
Qy 2476 SENGQITGVQOOTTHRNQAFMALEGQVITKKLHASIREKAGHW--FATTTPIIGKIMF 2533
Db 2469 WDPCKTILGIQELQKOLQNFISLDQLPMTPRYNDGRCLEGGKQPRFAAVPSVFGKIKF 2528
Qy 2534 AIKGRVTTGVSSTASEDSRKSASVSNAYYLDKMHYSIEGKDRTHYFVKIGSADGDLVTL 2593
Db 2529 AIKGRVTTGVSSTASEDSRKSASVSNAYYLDKMHYSIEGKDRTHYFVKIGSADGDLVTL 2588
Qy 2594 GTTIGRKVLESGVNVVTSQPTLLVNGRTRRPTNIEFOYSTLLSLTRYGLTDPDIDEKAR 2653
Db 2589 GNTGRRILENGVNVTSQPTLLVNGRTRRPTNIEFOYSTLLSLTRYGLTDPDIDEKAR 2645
Qy 2654 VLDQARQALATWAKAQKQARDREGSRRLTEGEKQOLLSTGRVQGVYEGYVILFVPEQYP 2713
Db 2646 VLEIARQARAQAWTEKQRRILQEGEGIRATEGEKQOLLSTGRVQGVYEGYVILFVPEQYP 2705
Qy 2714 ELADSSNIQFLRONEMGKR 2733
Db 2706 ELSDSANNIHFMRQSEIGRR 2725
RESULT 14
Q9P273 ID Q9P273 PRELIMINARY; PRY: 1769 AA.
AC Q9P273; 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE KIAA1455 PROTEIN (FRAGMENT).

GN KIAA1455.
OS Homo sapiens (Human).
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20277482; PubMed=10819331;
RA Nagase T., Kikuno R., Ishikawa K., Hirose M., Ohara O.;
RT "Prediction of the coding sequences of unidentified human
RT genes.XVII.The complete sequences of 100 new cDNA clones from brain
RT which code for large proteins in vitro.";
RL DNA Res. 7:143-150(2000).
DR EMBL; AB040888; BAA95979.1; .
FT NON_TER 1
FT NON_TER 1769 1769
SQ SEQUENCE 1769 AA; 199209 MW; 2E1524DC90F29125 CRC64;
Query Match 46.8%; Score 6822; DB 4; Length 1769;
Best Local Similarity 70.6%; Pred. No. 0;
Matches 1255; Conservative 255; Mismatches 248; Indels 20; Gaps 5;
Qy 833 QNSLLRCGRSDPLDIIOQG--QTDWPAVKSYDRILKLAGKDSHTIIPGPNFSSLSVL 890
Db 1 QNQPYCRGLPDPDIIISQSQSPQAQSYDRISFLIGSDSTHVIPIGESPFNKLASV 60
Qy 891 IRGOVVTTCGTPLVGNVSVKYPKYGTITRQDGTDLIANGASLTLLHFERAPFMSOE 950
Db 61 IRGOVLTADGTPGLVGNVSVFHYGYTITRQDGMFDLVANGASLTLLVFERPFLYQ 120
Qy 951 RTVWLPNWSFYAMDTLVKMTTEENIPSCDLSGFVRPDIISPLSTFFSAAPQGNPIV 1010
Db 121 HTVWLPNWSFYAMDTLVKMTTEENIPSCDLSGFVRPDIISPLSTFFSAAPQGNPIV 180
Qy 1011 ETQVLHIEIEPLGNSVRLRYLSSRTAGYKSLKLTMTQSTVPLNLRVHLVAVEGHLPQ 1070
Db 181 ETQVLHIEIEPLGNSVRLRYLSSRTAGYKSLKLTMTQSTVPLNLRVHLVAVEGHLPQ 240
Qy 1071 KSFOASPLASTFTWKTDAYQGVYGLSDAVSVGVGYETCPSLILWEKRTALQGLPEL 1130
Db 241 KWFASPLASTFTWKTDAYQGVYGLSDAVSVGVGYETCPSLILWEKRTALQGLPEL 300
Qy 1131 DPSNLGGWSLKHILNVKSGILHKGTCENQFLQQAIPATISMGNGRRRISCPSCNGL 1190
Db 301 DASNMGWTLDKHVLVDQNGILYKNGENQFISQPPVSSIMGNGRRRISCPSCNGL 360
Qy 1191 AEGNKLAPVALAVIDGSLVVGDFNYTRIRFPSPNVTISLELRNKEFKHNNPAHYL 1250
Db 361 ADGNKLAPVALAVIDGSLVVGDFNYTRIRFPSPNVTISLELRNKEFKHNNPAHYL 420
Qy 1251 AVDPVSGSLYSDTNSRRIRYRVKSLSGTKOLAGNSEVAVGTGEQCLPDEARCGDGGKAI 1310
Db 421 ATDPVTGDLVYSDTNTIRYRPSKLTGAKDLTKNAEVAVGTGEQCLPDEARCGDGGKAI 480
Qy 1311 DATLMSPRGIAVDKNGLMYFVDATMIRKVDQNGIISTLLGNSDLTAVRPLSCDSMVAQ 1370
Db 481 EATLMSPRGMAVDKNGLMYFVDATMIRKVDQNGIISTLLGNSDLTAVRPLSCDSMVAQ 540
Qy 1371 VRLWPTDLAVNPMDSIYVLENNVILRITENHQVSIITAGRPMHCQVPGIDISLSKLAIH 1430
Db 541 VRLWPTDLAVNPMDSIYVLENNVILRITENHQVSIITAGRPMHCQVPGIDISLSKLAIH 600
Qy 1431 SALESASAIASHTGVLVITETDEKKINLRQVTTNGEICILLAGAASDCCKNDVNCNY 1490
Db 601 TTLESATAIAVSYSGVLVITETDEKKINLRQVTTNGEICILLAGAASDCCKNDVNCNY 660
Qy 1491 SGDAYATDAIINSPSSSLAVAPDGTIYIADLGNIRIRAVSKNPKVNLAFNAFNAASPEQ 1550
Db 661 QSGDGYAKDAKLSAPSSLAASPDGTLXIADLGNIRIRAVSKNPKVNLAFNAFNAASPEQ 720
Qy 1551 ELYVFNADGIHOYVSVLTGELYVNFYSTNDVTELDNNGNSLKIIRDSGGRPHLLM 1610
Db 1551 ELYVFNADGIHOYVSVLTGELYVNFYSTNDVTELDNNGNSLKIIRDSGGRPHLLM 1610

Db	721	ELYIFDINGTHQYTSVLTGDIYNFYSNDNDITAVTDSNGNTLRIIRDPNRMPVRVVS	780
Qy	1611	PDNQIITLVGTNGLKVSTQNLZELGMLTYDQNTGLLATKSDTGWTFYDYDHEGRLT	1670
Db	781	PDNOVITLITGTNGLCKSMTAQGLVLFVTHGNSGLLATKSDTGWTFYDSEGRLT	840
Qy	1671	NVTRPGVVTSLHREMEKSIITDIENSRRDDVTVITNLSSEASVTVVODQVRNSYQLC	1730
Db	841	NVTFPGVVTNLHGDMKAITVDIESSEEDVSITNSLSDISFYTMVQDQLRNSYQIG	900
Qy	1731	NNGTLRVAVANGMISFHSSEPHVLACTITPTIGRCNISLPMENGLANSIEWLRKEQIKG	1790
Db	901	YDGLRIITVAGLSDHSIQEPHVLACTAMPTVAKRNNITLPGENGQNLVEMRFKEQAQK	960
Qy	1791	VTFGRKLRVHGRNLSIDYDRIRTEKIIDYDHRKFTLRIIYDQVRPELWLPSSGLAAV	1850
Db	961	VNVFGRKLRVGRNLSVDFRTTKTEKIIDYDHRKFTLRIIYDQVRPELWLPSSGLMAV	1020
Qy	1851	NVSYFFNGLRGLQGRAMSERDIDKQGRIVSRMFADGWKWSYSLDKSMVLLQSQRQY	1910
Db	1021	NVTSYTGITASIQRTTSEKVDYDQGRIVSRVFADGKTWSYTYLEKSMVLLHSQRQY	1080
Qy	1911	IFEYSSDRLLAVTSPVARHSMSTSTSIGYIRNIYNPPESNASVIFDYSDGRIILKTSF	1970
Db	1081	IFEYDMDRLSALTSPVARHTMOTIRSGYIRNIYNPPESNASITDNEGELLQTAFA	1140
Qy	1971	LGTGROVFKYKGLSKLSEIVDSTAVTFGYDETTGVLKMNVLQSGGFSCTIRYRKIGPL	2030
Db	1141	LGTSSRVLEKIRQRLSEILVDSTVSTYDETAGVLTAVNLQSDGFEITIRYRIGPL	1200
Qy	2031	VDKQIYRFESEGMVNAFEDYTHDNSFRIASIKPVISETPLPVDLYRYDEISGKVEHFK	2090
Db	1201	IDRQIFRSEDEGMVNAFIDYSY-DNSFRVTSMQGVINETPLPTDLQFDDISGKVEHFK	1259
Qy	2091	FGVIYDIQNIITAVMTLSKHFDTHGRIKEVOYEMFRSLMYMTVQYDSMGVRKREIKL	2150
Db	1260	FGVIYDIQNIITAVMTYTKHFDAGRIKEIQEYELFRSLMYMTIYQDNMGVRKREIK	1319
Qy	2151	LGPYANTYTYDYDGDGLOQSVAVNDRPTWRYSYDLNGLHLLNPGNSVRLMPLRYDLR	2210
Db	1320	IGPFANTYTYAYEYVDGLOQTLYLNEKIMWRYNYDLNGLHLLNPNNSARLTPLRYDLR	1379
Qy	2211	DRITRLGDVQYKTDGGLQSCGSDIFEYNSKGLLTRYNKASGWSVQYRYDGVGRASY	2270
Db	1380	DRITRLGDVQYRLDEGFLQRQGTETFEYSSKGLLTRYNSKSGSWTVIYRYDGLGRVSS	1439
Qy	2271	KTNLGHLLQYFYSDLNPNTRITHVYHNSSEITSLYDQGLHFLAMESSEGEYVVASDN	2330
Db	1440	KTSLGHOLOFFYADLTPTPTIRITHVYHNSSEITSLYDQGLHFLAMEISSGDEFIASDN	1499
Qy	2331	TGTPLAVFSLMILKQLOYTAYGEIYDSDNPDQFQVIGFHGGLYDPLTKLVHFTQRYD	2390
Db	1500	TGTPLAVFSSNGLMKQIOYTAYGEIYFSDNIDFQVIGFHGGLYDPLTKLVHFTQRYD	1559
Qy	2391	VLAGRWTSPDYTMKVNKGEPAPFNLYMFKSNPLSSELDLKNYTDVKSWSLWMEGFQLS	2450
Db	1560	ILAGRWTTPDIEIWKIRGKDPAPFNLYMFRNNPNASKIHVDKDIYDVSWSLWTFGFHLH	1619
Qy	2451	NIIPGPRAKMYFVPPPYELSESQASENGOLITGVQOQTTERRHNOAFMAL-EGGVITK	2506
Db	1620	NAIPGPPVPKFDLTPESEYELVSKQWDDIPPIFGVQOQVARQAKAFSLGKMAEQV---	1676
Qy	2507	KUHASTREKAGH-----WPAFTTPIIGKIMFAIREGRVTTGVSSITASDESKRVASVLNN	2561
Db	1677	-----SRRAGGAQSWLWTFATYKSLGKGMVLAWSQVRQTVNLNANEDCICKVAAVLNN	1731
Qy	2562	AVYLDKMHVYSIEGKDPHYFVKIGSADGLVLTGTTIGR	2599
Db	1732	ATYLENLHFTIEGKDPHYFVKIGSADGLVLTGTTIGR	1769

RESULT 15
Q9JLCO

ID	Q9JLCO	PRELIMINARY;	PRT;	930 AA.
AC	Q9JLCO;			
DT	01-OCT-2000 (TREMBlrel. 15, Created)			
DT	01-OCT-2000 (TREMBlrel. 15, Last sequence update)			
DT	01-DEC-2001 (TREMBlrel. 19, Last annotation update)			
DE	OD22 (FRAGMENT).			
GN	OD22.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Ben-Zur T., Feige E., Motro B., Wides R.;			
RT	"The mammalian Odz gene family: Homologs of a Drosophila pair rule			
RT	gene with expression implying distinct yet overlapping developmental			
RT	roles";			
RL	Dev. Biol. 117:107-120(2000).			
DR	EMBL; AF195419; AAF28317.1; -.			
FT	NON_TER 1			
SQ	SEQUENCE 930 AA; 106506 MW; 98879795182369DD CRC64;			
Query Match	33.2%;	Score 4848;	DB 11; Length 930;	
Best Local Similarity	98.9%;	Pred. No. 0;		
Matches 920;	Conservative 7;	Mismatches 3;	Indels 0;	Gaps 0;
Qy	1804	NLLSIDYDRNIRTEKIIDYDHRKFTLRIIYDQVRPELWLPSSGLAAVNSYFENGRLAGL	1863	
Db	1	NLLSIDYDRNIRTEKIIDYDHRKFTLRIIYDQVRPELWLPSSGLAAVNSYFENGRLAGL	60	
Qy	1864	QGRAMSERDIDKQGRIVSRMFADGWKWSYSLDKSMVLLQSQRQYIFEYDSSDRLLAV	1923	
Db	61	QGRAMSERDIDKQGRIVSRMFADGWKWSYSLDKSMVLLQSQRQYIFEYDSSDRLLAV	120	
Qy	1924	TPMSVARHSMSTSTSIGYIRNIYNPPESNASVIFDYSDGRIILKTSFDTGQVQVFKYK	1983	
Db	121	TPMSVARHSMSTSTSIGYIRNIYNPPESNASVIFDYSDGRIILKTSFDTGQVQVFKYK	180	
Qy	1984	LSKLSIIVYDSTAVTFGYDETTGVLKMNVLQSGGFSCTIRYRKIGPLVDKQIYRFESEGM	2043	
Db	181	LSKLSIIVYDSTAVTFGYDETTGVLKMNVLQSGGFSCTIRYRKIGPLVDKQIYRFESEGM	240	
Qy	2044	VNAREFTYTHDNSFRIASIKPVISETPLPVDLYRYDEISGKVEHFKGFGVIYDINQIIT	2103	
Db	241	INAREFTYTHDNSFRIASIKPVISETPLPVDLYRYDEISGKVEHFKGFGVIYDINQIIT	300	
Qy	2104	TAVMTLSKHFDTHGRIKEVOYEMFRSLMYMTVQYDSMGVRKREIKLGPYANTYTYD	2163	
Db	301	TAVMTLSKHFDTHGRIKEVOYEMFRSLMYMTVQYDSMGVRKREIKLGPYANTYTYD	360	
Qy	2164	YDGDGLOQSVAVNDRPTWRYSYDLNGLHLLNPGNSVRLMPLRYDLRDRITRLGDVQYKI	2223	
Db	361	YDGDGLOQSVAVNDRPTWRYSYDLNGLHLLNPGNSVRLMPLRYDLRDRITRLGDVQYKI	420	
Qy	2224	DDGGLQCQSGSDIFEYNSKGLLTRYNKASGWSVQYRYDGVGRRASVYKTNLGHLLQYFVS	2283	
Db	421	DDGGLQCQSGSDIFEYNSKGLLTRYNKASGWSVQYRYDGVGRRASVYKTNLGHLLQYFVS	480	
Qy	2284	DLNPNTRITHVYHNSSEITSLYDQGLHFLAMESSEGEYVVASDNTGTPLAIVESINGL	2343	
Db	481	DLNPNTRITHVYHNSSEITSLYDQGLHFLAMESSEGEYVVASDNTGTPLAIVESINGL	540	
Qy	2344	MIKQLQYTAIGEYIYDSDNPDQFQVIGFHGGLYDPLTKLVHFTQRYDQVYDLAGRWTSPDYTM	2403	
Db	541	MIKQLQYTAIGEYIYDSDNPDQFQVIGFHGGLYDPLTKLVHFTQRYDQVYDLAGRWTSPDYTM	600	
Qy	2404	WKNVGEKAPFNLYMFKSNPLSSELDLKNYTDVKSWSLWMEGFQLSNIIPGPRAKMYF	2463	
Db	601	WRNVGEKAPFNLYMFKSNPLSSELDLKNYTDVKSWSLWMEGFQLSNIIPGPRAKMYF	660	
Qy	2464	VPPPYELSESQASENGOLITGVQOQTTERRHNOAFMALEGVITKKLHASTREKAGHWFATT	2523	

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 18, 2002, 11:02:55 ; Search time 63.81 seconds
(without alignments)
4757.321 Million cell updates/sec

Title: US-09-800-198-8
Perfect score: 14581
Sequence: 1 MDVKDRRHRSITRGCGKEC.....ELADSSNIQFLRQNMGKR 2733

Scoring table: BLASTUM62
Gapop 10.0 , Gapext 0.5

747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802.*
1: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	14581	100.0	2733	22 AAU08680	Human FCTR3b polyp
2	14520.5	99.6	2724	22 AAU08681	Human FCTR3f polyp
3	6567	45.0	2136	22 AAM78695	Human protein SEQ
4	3915.5	26.9	2515	22 ABB71354	Drosophila melanog
5	3880.5	26.6	1045	22 AAB92858	Human protein sequ
6	3628.5	24.9	964	22 AAB93294	Human protein sequ
7	2837	19.5	1015	22 AAM79679	Human protein sequ
8	2335	16.0	625	22 AAB92780	Human protein sequ
9	2122.5	14.6	429	22 AAU00392	Human secreted pro
10	2022.5	13.9	746	22 AAG07028	Novel human diagno
11	1931	13.2	381	22 AAU08679	Human FCTR3a polyp

12	1931	13.2	381	22 AAB61141	Human NOV11 protei
13	1708.5	11.7	402	22 AAG04674	Novel human diagno
14	1250	8.6	242	22 AAG04673	Novel human diagno
15	1205	8.3	865	22 AAG03234	Novel human diagno
16	1203.5	8.3	768	19 AAW44817	Human gamma-heregu
17	1203.5	8.3	768	20 AAY06639	Gamma-heregulin
18	1203.5	8.3	768	21 AAY71176	Human heregulin va
19	1203.5	8.3	768	22 AAU09891	Human heregulin, g
20	1167	8.0	777	22 ABB65598	Drosophila melanog
21	1110	7.6	587	22 ABB59244	Drosophila melanog
22	1043.5	7.2	1124	22 ABB59241	Drosophila melanog
23	726	5.0	1810	17 AAR94563	Chicken cytotactin
24	713	4.9	136	22 AAG04672	Novel human diagno
25	680	4.7	2201	22 AAB36935	Human tenascin-C
26	668	4.6	527	19 AAW44818	Human gamma-heregu
27	666.5	4.6	2199	17 AAR94562	Human cytotactin
28	618	4.2	4618	22 AAM39043	Human polypeptide
29	596.5	4.1	214	22 AAB49657	Human SEC9 protein
30	560	3.8	151	21 AAB40417	Human ORFX ORF181
31	455	3.1	494	21 AAB25577	TIDE protein encod
32	455	3.1	567	21 AAB25595	protein encoded by
33	450.5	3.1	469	21 AAB25588	protein encoded by
34	444	3.0	1193	17 AAW05835	Chick Serrate. Ga
35	444	3.0	1193	21 AAY59599	protein encoded by
36	438.5	3.0	2321	19 AAW49698	Human Notch3 prote
37	434	3.0	636	22 AAB66269	Rat TANGO 272 SEQ
38	430.5	3.0	2471	20 AAY06816	Human Notch2 (humn
39	430	2.9	1010	20 AAW87896	Human JAGGED1 solu
40	430	2.9	1036	18 AAW18351	Proliferation and
41	430	2.9	1187	18 AAW18352	proliferation and
42	430	2.9	1208	19 AAW40827	Human Jagged prote
43	430	2.9	1218	17 AAW05833	Human Serrate-1 (H
44	430	2.9	1218	19 AAW44301	Human serrate 1.
45	430	2.9	1218	20 AAW87894	Human JAGGED1 prot

ALIGNMENTS

RESULT	1
AAU08680	
ID	AAU08680 standard; Protein; 2733 AA.
AC	AAU08680;
XX	
DT	18-DEC-2001 (first entry)
XX	
DE	Human FCTR3b polypeptide sequence.
XX	
KW	Human; FCTR3; myelogenous leukaemia; carcinoma; melanoma; glioma;
KW	astrocytoma; congenital neonatal alloimmune thrombocytopenia; infection;
KW	neurological disorder; neurodegenerative disorders; nerve trauma;
KW	familial myelodysplastic syndrome; Charcot-Marie-Tooth neuropathy;
KW	demyelinating Gardner syndrome; familial myelodysplastic syndrome;
KW	mental health condition; immunological disorder; allergy; infertility;
KW	bronchial asthma; Avellino type eosinophilia; lung disease; deafness;
KW	reproductive disorder; reproductive disorder; glycoprotein Ia deficiency;
KW	desmoid disease; turcot syndrome; liver cirrhosis; hepatitis C; virucide;
KW	gastric disorders; pancreatic disease; Schistosoma mansoni infection;
KW	Spinocerebellar ataxia; Plasmodium falciparum parasitaemia; diabetes;
KW	Corneal dystrophy-Greenow type I; Corneal dystrophy-lattice type I;
KW	Reis-Bucklers corneal dystrophy; cytostatic; immunosuppressive;
KW	antidiabetic; antidiabetic; antidiabetic; antidiabetic; antidiabetic;
KW	gynaecological; antidiabetic; antidiabetic; antidiabetic; antidiabetic;
XX	gene therapy; FCTR3b; neurestin-like protein.
OS	Homo sapiens.
XX	
PN	WO200166747-A2.
XX	
PD	13-SEP-2001.
XX	

PF 05-MAR-2001; 2001WO-US07160.
XX
PR 03-MAR-2000; 2000US-186592P.
PR 03-MAR-2000; 2000US-186718P.
PR 06-MAR-2000; 2000US-187293P.
PR 06-MAR-2000; 2000US-187294P.
PR 17-MAR-2000; 2000US-190400P.
PR 07-APR-2000; 2000US-196018P.
PR 03-JAN-2001; 2001US-259548P.
XX (CURA-) CURAGEN CORP.
XX
XX Vernet CAM, Fernandes E, Shimkets RA, Herrmann JL, Majumder K;
PI MacDougall J, Mishra V, Mezes PS, Rastelli L;
XX
XX WPI: 2001-596837/67.
DR N-PSDB: AAS14085.
XX
XX Novel polypeptides designated as FCTR polypeptides, useful in
PT detection, prevention and treatment of a broad range of pathological
PT states -
XX
XX Claim 1; Page 35-36; 215pp; English.
XX
XX The invention relates to human FCTR polypeptides, FCTR1-FCTR7, and the
CC nucleic acids encoding them. These sequences are useful for the treatment
CC or prevention of numerous disorders including myelogenous leukaemia,
CC carcinomas, melanomas, gliomas, astrocytomas, congenital neonatal
CC alloimmune thrombocytopenia, neurological disorders, neurodegenerative
CC disorders, nerve trauma, familial myelodysplastic syndrome,
CC Charcot-Marie-Tooth neuropathy, demyelinating Gardner syndrome, familial
CC myelodysplastic syndrome, mental health conditions, immunological
CC disorders, allergy and infection, bronchial asthma, Avelino type
CC eosinophilia, lung diseases, reproductive disorders, infertility, male
CC and female reproductive disorders, deafness, glycoprotein Ia deficiency,
CC desmoid disease, turcot syndrome, liver cirrhosis, hepatitis C, gastric
CC disorders, pancreatic diseases such as diabetes, Schistosoma mansoni
CC infection Spinocerebellar ataxia, Plasmodium falciparum parasitaemia,
CC Corneal dystrophy-Greenow type 1, Corneal dystrophy-lattice type I and
CC Reis-Bucklers corneal dystrophy. This sequence represents FCTR3b, a
CC neurestin-like protein.
XX
SQ Sequence 2733 AA;

Query Match 100.0%; Score 14581; DB 22; Length 2733;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2733; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MDVKRRHRLTRGRCGKECTRTSSLDSEDCRVPTQKSYSSSETLKAYDHDSDRMHNGR 60
|||||
1 mdvkdrrhrsltrgrcgkectrtssldsedcrvptqksssetlckaydhdsrmhngnr 60
|||||
61 VTDLTHRESDEPRGNTFTLAELGICPPSRHSGYCSDMGTLHGYSLSGSDSDSDTE 120
|||||
61 vcdllhresdeprgntftlaelgicpsprhsgycsdmgtlhggyslstsgdsdsdte 120
|||||
121 GGMSPHAIKRLGRIKSRSSGLSRNSALTLDSDNENKSDDENGRPTPTSPSLL 180
|||||
121 ggmshpahlrlwgrlksrsglsrnsaltldsdnenksddengrptptspssll 180
|||||
181 PSAQLPSSHNPPVSCQPLLDNSNTSHQIMDTNPDEEFPNSYLLRACSGPQASSGPP 240
|||||
181 psaqlpsshnppvcqplldnsntshqimdtnpdeefpnsyllracsgpqassgpp 240
|||||
241 NHHSSOSTLRPPLPPPHNTLSHHSSANSLSNRNSTLRNSQTHAPAPNDLATTPESVQ 300
|||||
241 nhhsostlrpplppphnhchshhssanslnrnslnrnsqthapapndlattpesvq 300
|||||
301 LQDSWVLSNVPLETRHFLKFTSSGSTPLFSSSSPGYPLTSGVTYTPPPRLPRNTFSRK 360
|||||
301 lqdswwlsvlnsvpletrhflkftssgstplfsssspgypltsqvtvytpprllprntfsrk 360
|||||

Qy 361 AFKLKPKSKYCSWKCAALSAIAAALLAILLAYFIVPWSLKNSSIDSGEAEVGRVTVQEV 420
|||||
Db 361 afklkpkskycswkcaalaaiaaallaillayfivpwslnssidsgaeavgrvtqev 420
|||||
Qy 421 PPGVFWRSQIHIHQPOFLKFNISLGKDALFGVYIRRGILPPSHAQYDFMERLDGKEKWSV 480
|||||
Db 421 ppgvfwrsqihisqpqlkfknislgkdalfgvyirrgilppshaqydfmerldgkekswv 480
|||||
Qy 481 ESPRRRSIQTLVQNEAVFVOYLDVGLWHLAFYNDGKEMVSFNTVVLDSVQDCPRNCH 540
|||||
Db 481 esprrrsiqtlvqneavfvoyldvglwhlafyndgkdkemvsfntvvlsvdqpcprnch 540
|||||
Qy 541 NGECVSGVCHCFPFGLGADCAKAAACPVLCSGNGGYSGKTCOCYSGWKGAECEDVPMNOCI 600
|||||
Db 541 ngecvsgvchcfpgflgadcaakaacpvlcsgnggysgktccycsgwkgaecdvpnmnci 600
|||||
Qy 601 DPSCGGHSCJIDGNCVCSAGYKGEHCEREVDCLDPTCSSHGVGVNGECLCSPGWGLNCEL 660
|||||
Db 601 dpscgghscjldgncvcvcsagykgheceevdcldpccshgvcvngecicspwggincel 660
|||||
Qy 661 ARVQCPDQCSGHGTLPDPTGLCSDPNMMGPDSCSVEVCSVDCGTHGVGIGACRCEEGWT 720
|||||
Db 661 arvqcpdqcsghgtlypdtglicsdpnmmgpdscsvevcsvdcgthgvigacrcceegwt 720
|||||
Qy 721 GAACQQRVCHPRCIEHGTCKDGKCECREGWNGEHCITIGRTAGTETDCGDLCLNGNGRGT 780
|||||
Db 721 gaacqqrvcprciehgtckdgkcecregwngehctigrtagtctdgcplcngngtrct 780
|||||
Qy 781 LGSNWCQVCOTGWRGPCNVMETSCADNKDNEGDLVDCLDPCCLQSAACONSLLCRG 840
|||||
Db 781 lgsnwcvcotgwrpgpcnvmetscadnkdnegdlvdcldpccclqsaacqnsllcrg 840
|||||
Qy 841 SRDPLDIIIOGQOTDMPAVKSFYDRILKLAGKDSHTIIPGENPFNSLSVLRGQVVTG 900
|||||
Db 841 srdpldiioqgotdmpavksfydrilklagkdsthiipgenpfnsllsvlrqgvvtg 900
|||||
Qy 901 TPLGVNVSVFKYPKYGYTIITRODGTDLIANGGASLTILHFERAFPMQERTVLPWNSF 960
|||||
Db 901 tplgvnsvfkykgytyitrogtfdlianggasltilhferafpmqertvlpwnsf 960
|||||
Qy 961 YAMDTLVNKTENSIPSCDLSGFVRPDIISPLSTFFSAAGQNPVTPETQVLEHIE 1020
|||||
Db 961 yamdtlvnmkteensipscdlsghfvrpdiisplstffsaagqnpvtpetqvlheie 1020
|||||
Qy 1021 LPSGNVKRLYSSRTAGYKSLKTKITWTOSTVPLNIRVHLMVAVEGHLFQKSFQASPLA 1080
|||||
Db 1021 lpsgnvkrlrysrtagysllktkitwtostvplnlrvhlmvaveghlfqksfqaspln 1080
|||||
Qy 1081 STFIWDKTDAYGQRYGLSDAVSVGYEYETCPSLILWEKRTALLQGFELDPSPNLGWSL 1140
|||||
Db 1081 stfiwdktdaygqryglsdavsvgyeyetcpslilwekrtallqgfeldpnsnlgws 1140
|||||
Qy 1141 DKHHLNVKSGILHKGTEGNOFTQOPAIITISIMGRRRRSISPCSNCLAEKNLLAPV 1200
|||||
Db 1141 dkhhlrvksgilhkgtegenqftqpaitsimgnrrrsiscpscnglaegnklapv 1200
|||||
Qy 1201 ALAVIDGSLVGDENVYIRRPFPSPNRVTSILELRNKEFHNNPAKYYLAVDPVSGSLY 1260
|||||
Db 1201 alavidgslvgydenvyirrfpsnrvtlsilelrnkefhnnpahkyylavdpvsgsly 1260
|||||
Qy 1261 VSDTNSRRIYRVKSLSGTKDLAGNSEVAGTCEQCLPDEARCGDGGKAIDATLMSPRGI 1320
|||||
Db 1261 vsdtnsrriyrvkslsgtkdlagnsevvagtceqclpdearcgddgkaidatlmprgi 1320
|||||
Qy 1321 AVDKNGLMYFVDATMIRKVDQNGIISTLLGSNDLTAVRPLSCDSSMDVAQVRLEWPTDLA 1380
|||||
Db 1321 avdknglmvfvdatmirkvdngiistllgsndltavrplscdssmdvavrlewptdla 1380
|||||
Qy 1381 VNPMDNSLYLVENNVLIRITENHQVSIITAGRPMHCQVPGIDYSLKLAHSALESASAIA 1440
|||||
Db 1381 vnpmdnslylvennvliritenhqvsiitagrpmhcqvpgidyslklaihalesasaia 1440
|||||
Qy 1441 ISHTGVLYITEDENKINRLRQVTNGETCLLAGAASDCDCKNDVNCYSGDDAYATDA 1500
|||||

Db 1441 ishtgviyifetdekklnrlrqvtntngeicllagaesdcckndvncncysgddaytada 1500
QY 1501 ILNPSLAVAPDGTIYIADLGNIRIRAVSKNKPVLNAFYEAASPGOEYLYVFNADGI 1560
Db 1501 ilnpslslavapdgltiyadlgniriravsknkpvlnafqyeaspgqelyyfnadg1 1560
QY 1561 HQTVSLVTGEYLYNFYSTDNDVTELIDNNGNSLKIIRDSGMPRHLLPMDNQIITLV 1620
Db 1561 hqtvslvtgeylynfystdndvteididnngnslkrrdssgmprhlmpdnqiliclv 1620
QY 1621 GTNGLKVVSTONLEGLMYYDGNLTGATKSDETGWTTFYDYDHEGRLNVRPTGVVT 1680
Db 1621 gtngglkvstqnllelmydydngntllatksetgwttfydydhegrltnvtrptgvvt 1680
QY 1681 SLHREMEKSTIDIENSRRDDVTITNLSVSEASYTVVODQVRNSYOLCNGNGLTRVMA 1740
Db 1681 slhremekstidlensrrddvtitnlsveasytvdqdvznsyqlcnnngclrvmya 1740
1741 NGMISPHSEPHVLAGTITPTIGRCNISLPMENGLNSIEWRLRKEQIKGVITFGRKLRY 1800
Db 1741 ngmishsephvlagtittptigrncnislpmenglnsiewrlrkeqikgvitfgrklrv 1800
QY 1801 HGRNLLSIDYDRNIRTEKIIYDDHRKFTLRIYDOVGPPFLWLPSSGLAANVSYFFNGRL 1860
Db 1801 hgrnllsidydrnirtekiyddhrkftlriydgvgprflwlpssglaavnvsvffngrl 1860
QY 1861 AGLQGMASERTIDKGRIVSRMFADGKVMYSYLDKSMVLLLOSQRQYIFEVDSDDL 1920
Db 1861 aglqgmaseriddkgrivsrmfadgkvmysyldksmvlllqsqrqyifeydsddl 1920
QY 1921 LAVTMPVARHSTHTSISYIRNIYNPPESNASVIFDYDDGRILKTSFLGTGRQVYFK 1980
Db 1921 lavtmpvarhsthtsisyirniynppesnasvifdydsddgrilktstflgtgrqvyfk 1980
QY 1981 YGKLSKLSIYVYSTAVTFGDETTGVLMKNVLOSQGFSTIRYRKIGPLVDQIYRFE 2040
Db 1981 ygklsklsiyvystavtfgdetgtvlmknvlosqgfstiryrykigplvdqiyrfse 2040
QY 2041 EGMVNARFDYTHDNSPRIASIKPVISETPLPVDLYRYDEISGKVEHFGFVYIYDINO 2100
Db 2041 egmvnarfdythdnspriasikpvisetplpvdlyrydeisgkvehfgfvgviyding 2100
QY 2101 IITAVMTLSKHEDTHGRIKEVOYEMFSLMYMTVOYDSNGRVIKRELKLGYPYANTTKY 2160
Db 2101 iitavmtlskhfthgrikevqyemfslmymtvqydsngrvikrelklgpyanttky 2160
2161 TYDYGQGLQSVAVNDRPTWRYSYDLNGLNLLNPGNSVRLMPLRYDLDRITRLGDVQ 2220
Db 2161 tydydgqglqsvavndrptwrysdydingnlhllnpgnsvrlmplrydlrditrlgdvq 2220
QY 2221 YKIDDDGLCQSGSDIFEYNSKGLLTRYANKASGWSVOYRYDVGRRASYKTNLGHHLQY 2280
Db 2221 ykiddyglcrgsdifeynskgltryankasgwsvoirydygrrasyktnlghhlqy 2280
QY 2281 FYSOLHNPTRIITHYHNSNSEITSLYDLOGLHFAMESSSGEEYVYASDNTGTPLAYFSI 2340
Db 2281 fysolhnptriithynsnseitslyldlgghlfamesssgeeyvasdntgtplavfsi 2340
QY 2341 NGLMIKQLOYTAGEIYYDSNPDMQWIGPHGGLYDPLTKLVHFTQRDQDVLAGRWTSPPD 2400
Db 2341 nglmikqlqytageiyydsnpdmqwigphgglydpltklvhftqrdqdvlagrwtsppd 2400
QY 2401 YTMKNKVGKEPAPNLYMFKSNPNLSSELDELKNVYTDVDSKWLVMFGFQLSNIIPGFPRAK 2460
Db 2401 ytmknkvgepapnlymfksnnpnlsseldelknyvtdvdskwlvvmfgfqlsnlpgfpprak 2460
QY 2461 MYFVPPPYELSESQASENGQILITGVQQTTERHQAFALESGQVITKHLHASIREKAGHWF 2520
Db 2461 myfvpypyelseqasengqiligtvqqtterhqaafalesgqvltkhlhasirekaghwf 2520
QY 2521 ATPTPIITGKIMFAIKGGRVTGTVSSIASEDSRKVASVLNNAYYLDKMHYSIEGKQTHYF 2580
|||||

2521 atttptligkimfaikgrvtgtvssiasedsrksvaslnnayyldkmhysieqkdthyf 2580
QY 2581 VKIGSADGDVLTGTTIGRKVLESGVNVTVSQPTLLVNGTRRRTNTEFOYSTLLLSIRY 2640
|||||
Db 2581 vkigsadgdvltglttigrkvlesgvnvtvsqptllvngtrtrfnlefystlllsiry 2640
QY 2641 GLTPDTLDEBKARVLDQARALGTAWAKEQQKARDREGSRWLWTEGEKQOLJSTGRVQG 2700
Db 2641 gltpdtldeekarvldqaralgtawakeqqkardregsrwlwtegeqkqllstgrvqg 2700
QY 2701 YEGYVYLPVQYQPELADSSSNIOFLRONEMGKR 2733
|||||
Db 2701 yegyvyvpveqypeladsssnlqflrqnmgkr 2733
RESULT 2
AAU08681
ID AAU08681 standard; Protein; 2724 AA.
XX AC AAU08681;
XX AC
XX DT 18-DEC-2001 (first entry)
XX DE Human FCTR3f polypeptide sequence.
XX DE
KW Human; FCTRX; myelogenous leukaemia; carcinoma; melanoma; glioma;
KW astrocytoma; congenital neonatal alloimmune thrombocytopaenia; infection;
KW neurological disorder; neurodegenerative disorders; nerve trauma;
KW familial myelodysplastic syndrome; Charcot-Marie-Tooth neuropathy;
KW demyelinating Gardner syndrome; familial myelodysplastic syndrome;
KW mental health condition; immunological disorder; allergy; infertility;
KW bronchial asthma; Avellino type eosinophilia; lung disease; deafness;
KW reproductive disorder; reproductive disorder; glycoprotein Ia deficiency;
KW desmoid disease; turoct syndrome; liver cirrhosis; hepatitis C; virucide;
KW gastric disorders; pancreatic disease; Schistosoma mansoni infection;
KW Spinocerebellar ataxia; Plasmodium falciparum parasitaemia; diabetes;
KW Corneal dystrophy-Greenow type I; Corneal dystrophy-lattice type I;
KW Reis-Bucklers corneal dystrophy; cytostatic; immunosuppressive;
KW anti-allergic; antilasthmatic; antifertility; antiinflammatory;
KW antidiabetic; protozoacide; hepatotropic; virucide; ophthalmological;
KW gynaecological; antifertility; immunostimulant; auditory; haemostatic;
KW gene therapy; FCTR3a; neurestin-like protein; FCTR3f.
XX OS Homo sapiens.
XX OS
XX PN WO200166747-A2.
XX PF
XX PD 13-SEP-2001.
XX PF
XX PF 05-MAR-2001; 2001WO-US07160.
XX PR 03-MAR-2000; 2000US-186592P.
XX PR 03-MAR-2000; 2000US-186718P.
XX PR 06-MAR-2000; 2000US-187293P.
XX PR 06-MAR-2000; 2000US-187294P.
XX PR 17-MAR-2000; 2000US-190400P.
XX PR 07-APR-2000; 2000US-196018P.
XX PR 03-JAN-2001; 2001US-259548P.
XX PA (CURA-) CURAGEN CORP.
XX PA
XX PI Vernet CAM, Fernandes E, Shimkets RA, Herrmann JL, Majumder K;
PI Macdougall J, Mishra V, Mezes PS, Rastelli L;
XX PI
XX XX WPI; 2001-596837/67.
DR N-PSDB; AAS14089.
XX XX
PT Novel polypeptides designated as FCTRX polypeptides, useful in
PT detection, prevention and treatment of a broad range of pathological
XX states -
XX Claim 1; Page 39; 215pp; English.
XX XX

CC The invention relates to human FCTR3 polypeptides, FCTR1-FCTR7, and the
CC nucleic acids encoding them. These sequences are useful for the treatment
CC or prevention of numerous disorders including myelogenous leukaemia,
CC carcinomas, melanomas, gliomas, astrocytomas, congenital neonatal
CC autoimmune thrombocytopaenia, neurological disorders, neurodegenerative
CC disorders, nerve trauma, familial myelodysplastic syndrome,
CC Charcot-Marie-Tooth neuropathy, demyelinating Gardner syndrome, familial
CC myelodysplastic syndrome, mental health conditions, immunological
CC disorders, allergy and infection, bronchial asthma, Avellino type
CC eosinophilia, lung diseases, reproductive disorders, infertility, male
CC and female reproductive disorders, deafness, glycoprotein Ia deficiency,
CC desmoid disease, turocot syndrome, liver cirrhosis, hepatitis C, gastric
CC disorders, pancreatic diseases such as diabetes, Schistosoma mansoni
CC infection Spinochorellar ataxia, Plasmodium falciparum parasitaemia,
CC Corneal dystrophy-Greenow type I, Corneal dystrophy-lattice type I and
CC Reis-Bucklers corneal dystrophy. This sequence represents FCTR3f, a
CC homologue of FCTR3a protein.
XX
SQ Sequence 2724 AA;

Query Match 99.6%; Score 14520.5; DB 22; Length 2724;
Fast Local Similarity 99.7%; Pred. No. 0;
Matches 2724; Conservative 0; Mismatches 0; Indels 9; Gaps 1;

QY 1 MDVKRRHRLRGRGKRECVTSSLSDEDCRVPTQKSYSSSETLKAYDHDHSMHYGNR 60
DB 1 mdvkrhrslrgrcgkrcvrtsslsdsedcrrvptqksssetlkdahdhsrmhygnr 60
QY 61 VDTLTHRESDEPRGNTFTLAELGICEPSPHRSYGCSDMGILHOGYSLSGSDADSDE 120
DB 61 vdtlthresdeprgntftlaelgicepsphrsgydcsmgllhogsystsgdsadsdce 120
QY 121 GGMSPHAIIRLGRGKRRSSGLSSRENSALTLDSDNENKSDDENGRPIPTSPSLL 180
DB 121 ggmspehaiirlgrgkrrssglsrensaltldsdnenksddengrpipptspssll 180
QY 181 PSAQLPSSNHPVSCQPLLDNSNTSHQIMTNDPDEEFPNSYLLRACSGPOQASSGPP 240
DB 181 psaqlpssnhppvscqplldnsntshqimndpdeefpnsyllracsgpqgassgpp 240
QY 241 NHHQSSTLRPLPPPHNHTLSHHSSANSLSNLSLTNRRSQTHAPAPNDLATTPESVQ 300
DB 241 nhhsqstlrplppphnhtlshhssanslnsltnrrsqthapapndlattpesvq 300
QY 301 LODSWLNSVPLETRHFLFKTSGSTPLFSSSPGYPLTSGVTYTPPPRLPLRNTFSRK 360
DB 301 lqdswwlnsnvpletrhflfktsgstplfsssspgypltsqvtvytppprlprntfsrk 360
QY 361 AFKLKPKSKYCKWKAALSAIAAALLAILLAYFVPSLKNSSSDSGEAEVGRRTQEV 420
DB 361 afklkpkskycwkaalasaiaaallailayfvpwslknssidsgeaevgrvtqev 420
QY 421 PPGVFWRSQIHISQPFKFNLSLGDALFGVYIRRGLPSPSHAQYDEMERLDGKESVY 480
DB 421 ppgvfwrsqihisqpfkfnlsldgalfgvyrirrglpsshaqydfmerldgkewsvv 480
QY 481 ESPRRRSITQLVQNEAVFVQYLDVGLWHLAFYNDGKEMVSFNTVLDSDQDCPRNCH 540
DB 481 esprrrsrltqlvqneavfvqyldvglwhlafyndgkkmvsnfntvldsvdqpcrnch 540
QY 541 GNGECVGVCHCFPGFLGADCAKACPVLCSSNGQYSGTKCYCGWKGAECDVPMNQI 600
DB 541 gngecvsgvchcfpgflgadcaakacpvlcsgngqysgkcygkwgaecdvpnmnqi 600
QY 601 DPCSGHGSCIDGNCVCSAGYKGEHCEEVDCUDPTCSSHGVGVNCECLCSPGMGLNCEL 660
DB 601 dpcsgghgscidgncvcasagykgeheevdclptcsshgvgvnecelcspgwglncel 660
QY 661 ARVQCPDQCSGHGTLYLPTDGLSCDPNMWGPCSVCEVCSVDCGTHGVCIGACRCEGWT 720
DB 661 arvqcpdqcsghgtlylptdglscdpnmwgpdcsvcevcsvdcgthgvcigacrceegt 720

QY 721 GAACDQVRCHPRCIEHGCTKDKGCECREGNGEHCTIGQTAGTETDCCPDLCNGNRC 780
DB 721 gaacdqrchprciehgtckdkgkcecregnghehcti-----dgcpcdlcngnrc 771
QY 781 LGQNSQOCVQGWGPGCGCNVAMETSCADKNDEGDLVDCLDPPCCLOSACONSLLCRG 840
DB 772 lqgnswqvcvqgwpgcgvnametscadknknegdlvdcldppccclqscqcnallcr 831
QY 841 SRDPLDIIOQGGTDMPAVKSFYDRIKLAGKDSHTIIPGENPFSLSVSLIRGQVVTIDG 900
DB 832 srpdldiioqggtwdmpavksfydrliklagkdsthiipgenpfnselvalirgqvvtidg 891
QY 901 TELVGVNSFVYKPYKTYTTRQDQTFDLIANGASLTLHFERAPFMSQERTVWLPWNSF 960
DB 892 tplvgvnsvfvykpyktyttrqdgtdfdliangasltlhferapfmsqertvwpwnsf 951
QY 961 YAMDTLVKMTKEENSIPSCDLSGFVRDPDPIIISPLSTFSAAPGONPIVPTQVLEHETE 1020
DB 952 yamdtlvkmkteensipscdlsghfvrpdpilissplstfssaapgnpivpqcvlhee 1011
QY 1021 LFGSNVKLYRLSSRTAGYKSLIKITMTQSTVPLNLRVHLMVAVEGHLFOKFSQASPL 1080
DB 1012 lpgsnvklrlylssrtagyksllkitmtqstvpnlrvhlmvaveghlfqkfsqaspl 1071
QY 1081 STFIWKTDAYQORYGLSDAVVSVGFYETCPSLILHKEKRTALLQGPFLDPSNLGGWSL 1140
DB 1072 stfiwktdayqoryglsdavvsvgfeyetcpslilwekrtallqgfaldpsnlggws 1131
QY 1141 DKHHILNVKSGTLHKGTTGENQFLTQOPAITSIMNGRRRSISCPSCNGLAGNKLLAPV 1200
DB 1132 dkhhilnvksgtlhkgttgenqfltqgpaitsimngrrrsiscpscnglaegnkllapv 1191
QY 1201 ALAVGIDSLYVGDFNYIRRIFFPSRNVTSLILRNKEFKHNNPAHKYYLAVDPVSGSY 1260
DB 1192 alavgidsglyvgdfnyirrifpsrnvtslilrnkefkhsnnpahkyylavdpvsgsy 1251
QY 1261 VSDTNSRRIRYRVKSLSGTKDLAGNSFVAGTGEQCLPDEARCGDGGKAIDATLMSPRGI 1320
DB 1252 vdsdtnsriryrvkslsgtkdlagnsfvagtgeqclpdearcgdggkaidatlmprgi 1311
QY 1321 AVDKNGLMYFVDATMIRKVDQNGIISTLLGSNDLTAVRPLSCDSSMDVAQVLEWPTDL 1380
DB 1312 avdknglmyfvdatmirkvdngiistllgsgndltavrplescdssmdvaqvlewptdl 1371
QY 1381 VNPMDNSLYVLENNVILRITENHQSIIAGRPWHQVPGIDYSLSKLAIHSALESASATA 1440
DB 1372 vnpmdnslyvlennvilritenhqsiiagrpwhqvpgidyslsklaahsaaesata 1431
QY 1441 ISHTGVLYTETDEKKINRLROVTTNGETCLLAGAASDCKDCKNVCNCSGDDAYATDA 1500
DB 1432 ishtgvlyteteckkinrlrovttngetcllagasodckdcknvcncysgddayatda 1491
QY 1501 ILNPSSSLAVAPDGTIYIADLGNIRAVSKNPNVNAFNOYEAASPGQEULYVFNADGI 1560
DB 1492 ilnpssslavapdgtiyiadlgniravsknpvlnafnoyeeaaaspgqeulyvfnadgi 1551
QY 1561 HOYTYSVTGEYLYNFTYSTDNDVTTELIDNNGNSLKIRDDSSGMPRHLLMPNQIITLV 1620
DB 1552 hqytsvltgeylynfytystdndvttelidnngnslkirddssgmprhllmpdnqitltv 1611
QY 1621 GTNGLGVVSTQNLGLMTYDGTNGLLATKSDGTWTFYDYDHEGRLTNVTRTGVVT 1680
DB 1612 gtnnglgvvstqnlglmtydgtngllatksdgtwtfydydhegrltnvtrtgvvt 1671
QY 1681 SLHREMEXSITIDIENSRRDDVTITNLSSVEASVTVQDVNRNSYQLCNGNGLRVMYA 1740
DB 1672 slhremeksitidienrrddvtitnlssveasvtyvqdvnrnsyqlcngntrlrmya 1731
QY 1741 NGMGISFHSHEPHVLGAGTTPTIGRCNLSLPMENGLNSLEWRLEKQIKGVKTFGRKLVR 1800
DB 1732 ngmgisfhshephvlgagttptigrclnslpmenglnslewrleqikgvktrfgrklvr 1791
QY 1801 HGRNLLSIDYDRNIRTEKIYDDHRRKFTLRIIYDQVGRFPLMLPSSGLAANVSYFNGRL 1860

QY 1866 GAMSERTDIDKOGRIVRMPADGKWSYSYLDKSMVLLLSQSQRYIFEYDSSDRLLAVTM 1925
Db 181 gttsekvdvqggqrvsarvfdagktwaytyleksmvlllshsqryifeymwdrleaitm 240
QY 1926 PSVARHSMSTHTSGYIRNLYNPESNASVIFDYSDGRILKTSFLGTGRQVFKYKGLS 1985
Db 241 psvarhsmqctiralggyrnllynpesnaslftdyneeglllqltaflgtsrvlfkryrt 300
QY 1986 KLSEIVDSTAVTFGDTGVGLKMWNLQSGFSGTIRYRKIGPLVDQKIYRFSEEGMVN 2045
Db 301 rleailydstvstydtagvktvnlqsdgfictiryqipldrqkfrfseidgmvn 360
QY 2046 ARDPTVYHDSFRIASIKPVISETPLVDLYRYDEISGKVEHFGKFGVVIYDNIQTTA 2105
Db 361 arldysy-dnsfrvsmqgvlnetplidlyfdidsgkveqfgkfgvlyvdydnqlaist 419
QY 2106 VMTLSKHFDTHGRKEVOYEMFSLMWTVOYDSDMGRVTKRELKGLGYPYANTTKTYTDYD 2165
Db 420 vmytkhfdahgrikeiqeifrslymwticqydmgrvrtkreklgpfanttkyayeyd 479
QY 2166 GDGQLOSVAVNDRTWRYSDNLGNLHLLNPGNSVRLMPLRYDLRITRLGDVQYKIDD 2225
Db 480 vdggqlqvtvynelkmwryndngnlhlnpsnsarilptryldrdrtrtldgvdqyrlde 539
QY 2226 DGVLCQRGSDIFPYNKGLTRAYNKASGWSVQVRYDGVGRRASYKTNLGHLLQYFYSDL 2285
Db 540 dgflrqtgelfeysskglltrvysksgwtviyrydglrrvssktslghlqfyadl 599
QY 2286 HNPTRITHVNHNSSETLSLYDLOGLHFAWESSGVEYVVASDNTGTPLAVERINGLMI 2345
Db 600 tpyrlichvynhasseltslydqlghlfameissgdeyfasdntgtplavfssnglml 659
QY 2346 KQLQYTAAGYIYDSNPDQFQWGFHGLGLYDPLTKLVHFTQRQYDVLVLAGRWTSPTDTMWK 2405
Db 660 kqigytaygeiyfdsnldfvlqfghgllydpltklihfgedydylagrtwtpdieiwk 719
QY 2406 NVGKEPAPFNLYWFKSNPNLSSELDLKNYVTDVKSMLVMFGFQLSNLIIPGPRAKMYFVP 2465
Db 720 rikdpapfnlymfrnnpaskihdvkdytdvnswlvtfgfhlhnaipgfpvpkfdte 779
QY 2466 PPSYELSESQASENGOLITGVQOQTERHNOAFMAL-EGQVITTKKLHASIREKAGH- 2518
Db 780 pseyelksqqwdlplpfgvqqvraqakafslgkmaevqv-----srrraggaqs 831
QY 2519 --WFATPTPIIGKIMFAIKGRTVTGVSSIASDSRKVASVLNNAYLLDKMHSYIEGKD 2576
Db 832 wlfatvksligkvmlavsggrvqtnvlnianedcikaavlnnafylenlhftiegd 891
QY 2577 THYFVKIGSADGLVLTGTTIGRKVLESGVNVVVSQPTLLVNGRTRFRFTNIEFQYSTLL 2636
Db 892 thyfikttdpesdgltrltsgskalenglnvtvsgattvngtrrrfadvemqfgalal 951
QY 2637 SIRYGLTPDPTLDEKARVLQDQARALGTAWAKEQKARDGREGSLRWTEGEKQOLLSTG 2696
Db 952 hvrygm---tdcekarileqarqlarawareqqrdrdgegarlwtgeqxrqlisag 1008
QY 2697 RVQYEGYVLPVEQYQPELADSSNIOFLRQNMGR 2733
Db 1009 kvagdygyvlsveqypeladsannqlrqsgeigr 1045
RESULT 6
AAB93294
ID AAB93294 standard; Protein; 964 AA.
XX
AC AAB93294;
DT
XX 26-JUN-2001 (first entry)
DE Human protein sequence SEQ ID NO:12355.
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.
KW
XX

OS Homo sapiens.
PN EP1074617-A2.
XX
PD 07-FEB-2001.
PF
XX 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602
full-length cDNAs defined in the specification, and for the detection
and/or diagnosis of the abnormality of the proteins encoded by the
full-length cDNAs -
PS Claim 8; SEQ ID 12355; 2537pp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesizing 5602
full-length cDNAs defined in the specification. Where a primer set
comprises: (a) an oligo-dT primer and an oligonucleotide complementary
to the complementary strand of a polynucleotide which comprises one of
the 5602 nucleotide sequences defined in the specification, where the
oligonucleotide comprises at least 15 nucleotides; or (b) a combination
of an oligonucleotide comprising a sequence complementary to the
complementary strand of a polynucleotide which comprises a 5'-end
sequence and an oligonucleotide comprising a sequence complementary to a
polynucleotide which comprises a 3'-end sequence, where the
oligonucleotide comprises at least 15 nucleotides and the combination of
the 5'-end sequence/3'-end sequence is selected from those defined in
the specification. The primer sets can be used in antisense therapy and
in gene therapy. The primers are useful for synthesizing polynucleotides,
particularly full-length cDNAs. The primers are also useful for the
detection and/or diagnosis of the abnormality of the proteins encoded by
the full-length cDNAs. The primers allow obtaining of the full-length
cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
represent oligonucleotides, all of which are used in the exemplification
of the present invention.
XX Sequence 964 AA;
QY Query Match 24.9%; Score 3628.5; DB 22; Length 964;
Best Local Similarity 69.0%; Pred. No. 9.4e-194;
Matches 673; Conservative 139; Mismatches 143; Indels 21; Gaps 5;
QY 1767 ISLPMENGLNSIEWLRKEQIKGVITFGKRLVHGNLLSIDYDRNIRTEKIYDDHRKF 1826
Db 1 mtlpgngqnlvewfrfkeqagkvngvgrklrvngnllsvdfrttkktkiyddhrkf 60
QY 1827 TLRIYDQVGRPFLWLPLSSGLAAVNVSYFFNGRLAGLQAGMSERTDIDKQGRIVSRMFA 1886
Db 61 llriaydtsghptlwiwpskklmavnvtystgqiaslgrgttsekvdvdydgggrivsvrfa 120
QY 1887 DGKWSYSYLDKSMVLLLSQSQRYIFEYDSSDRLLAVTMPSVARHSMSTHTSGYIRNII 1946
Db 121 dgkwtstytleksmvlllshsqryifeymwdrleaitmptsvartmqtirsiggyrniy 180
QY 1947 NPESNASVIFDYSDGRILKTSFLGTGRQVFKYKGLSKLSLSEIVDSTAVTFGDETTG 2006
Db 181 npesnaslftdyneeglllqltaflgtsrvlfkryrtldstsvrsvtydtag 240

QY	2007	VLKMYNLQSGFSCTIRYKKGIPGVKQIYRSEBGMVNAERDYYTHDNSFRIASIKPVI	2066
Db	241	vlktnvlnqsdgfiictiryqglpdlrqlfrfsebgmgnarfdisy-dnsfrvtsmqgvi	299
QY	2067	SETPLVDLYRYDETSKGVHFGKGVYIYDINQIITTAVMTLKSHFDTHGRKEIQVDEM	2126
Db	300	netplidlyqfddlsqgvegfkgfgyvdyngdqnglqistavmtytkhfdahgrikelgyei	359
QY	2127	FRSLMYMTVOYDSMGVYKRLKGLGPVANTTKYNYDVGQQLQSVAVNDRPTWRYSD	2186
Db	360	fslmywtlqymngvrtckreiklpgfantckeyeydvqglqcyvlneximwrynyd	419
QY	2187	LGNLHLLNPGNSVRLMLPLRYDLRDLRITRLGLDVQYKIDDDGYLCQSGDIEFYNKSGLLT	2246
Db	420	lgnlhllnpnsarltplydlrldrlrclgqvrylrdedglrqtgeifeysskgilt	479
QY	2247	RAYNKASGWSVOYRVDGVRASRYKTNLGHLOYPFSLNPTTRTHVYNINSNITSILY	2306
Db	480	rvysksgwtviyrydgyigrrvssktslqhlqfyyadltyptrlthvynhsseitsly	539
QY	2307	YDQLGHLFAMESSEGEYVVASDNTGTPLAVFSINGLMKLOLYTAYGEIYYDSNPDQOM	2366
Db	540	ydlqqlhlfamelssgedefiasdntgtplavfssnglmkqiqtaygeiyfidsnidfql	599
QY	2367	VIGFHGGYDPLTKLVHFTQRDYDVLAGRWTSPDYTMKNVKGEPAPFNLYMFKSNPNLS	2426
Db	600	vlgfhgglydpltklhfgerdydilagrwtptdplkwkrgkdpapfnlymfrnnpas	659
QY	2427	SELDLKNVYTDVKSVMWFGQLSNIIIGFPRAKMYFVPPPYELSESQASENGQILITGVQ	2486
Db	660	kldvkdvtvsnslvltfghlnaipfpvpkfdltpsvelvksqwdidpifgvq	719
QY	2487	QOTTERHNOAFNAL----EQGVTKKLHASIREKAGH-----WFATTPTIIGKGINFAIKE	2537
Db	720	qvarqakafislgkmaevq-----srraagagswlfatvkslkgvmlavsq	771
QY	2538	GRVTTGVSSIASERDRKVASVUNNAYYLDKMHYSTEGDTHFYFKIGSADGDLVTLGTTI	2597
Db	772	grvqtnvlnianedekvvaavlnnafylenhftlegkdkthfyfkttpesdlglrlts	831
QY	2598	GRKVLSEGVNVTVSOPTLLVNGTRRFTNIEFOYSTLLSIRYGLTPDTLDEEKARVLDQ	2657
Db	832	grkalenglnvtcsdttvngtrfrfadvmqfgalalhrygm---tldeekarileq	888
QY	2658	ARQALGTAWAKEQKARDREGSRSLWTEGEKQQLLSTGRVQGYEGYVLPVEQYPELAD	2717
Db	889	arqalarawareqrvrldgegarlwtegekrqllsagkvgydygyvlsveqypelad	948
QY	2718	SSSNTQFLRQNMGRK 2733	
Db	949	sannlqflrqseigr 964	
RESULT	7		
AAW79679			
ID	AAW79679	standard; Protein; 1015 AA.	
XX	AC	AAW79679;	
XX	AC	AAW79679;	
DT	06-NOV-2001	(first entry)	
DE	Human	protein SEQ ID NO 3325.	
KW	Human;	cytokine; cell proliferation; cell differentiation; gene therapy;	
KW	vaccine;	peptide therapy; stem cell growth factor; haematopoiesis;	
KW	tissue growth factor;	immunomodulatory; cancer; leukaemia;	
KW	nervous system disorder;	arthritis; inflammation.	
OS	Homo sapiens.		
PN	W0200157190-A2.		
XX			

PD	09-AUG-2001.		
XX	05-FEB-2001; 2001WO-US04098.		
XX	03-FEB-2000; 2000US-0496914.		
PR	27-APR-2000; 2000US-0560875.		
PR	20-JUN-2000; 2000US-0598075.		
PR	19-JUL-2000; 2000US-0620325.		
PR	01-SEP-2000; 2000US-0654936.		
PR	15-SEP-2000; 2000US-0663561.		
PR	20-OCT-2000; 2000US-0693325.		
PR	30-NOV-2000; 2000US-0728422.		
XX	(HYSE-) HYSEQ INC.		
PA	Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;		
PI	Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;		
PI	Xue AJ, Yang Y, Wejhrman T, Goodrich R;		
XX	WPI; 2001-476283/51.		
DR	N-PSDB; AAK52812.		
XX	Nucleic acids encoding polypeptides with cytokine-like activities,		
PT	useful in diagnosis and gene therapy -		
XX	Claim 20; Page 307; 6221pp; English.		
XX	The invention relates to polynucleotides (AAK51456-AAK53435) and the		
CC	encoded polypeptides (AAM78323-AAK80302) that exhibit activity elating to		
CC	cytokine, cell proliferation or cell differentiation or which may induce		
CC	production of other cytokines in other cell populations. The		
CC	polynucleotides and polypeptides are useful in gene therapy, vaccines or		
CC	peptide therapy. The polypeptides have various cytokine-like activities,		
CC	e.g. stem cell growth factor activity, haematopoiesis regulating		
CC	activity, tissue growth factor activity, immunomodulatory activity and		
CC	activation/inhibin activity and may be useful in the diagnosis and/or		
CC	treatment of cancer, leukaemia, nervous system disorders, arthritis and		
CC	inflammation.		
CC	Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666		
CC	(AAM80020) are omitted as the relevant pages from the sequence listing		
CC	were missing at the time of publication.		
XX	Sequence 1015 AA;		
SQ			
Query Match	19.5%; Score 2837; DB 22; Length 1015;		
Best Local Similarity	53.6%; Pred. No. 1.4e-149;		
Matches	542; Conservative 191; Mismatches 267; Indels 12; Gaps 5;		
QY	1707	TNLSVEASYTVVQ--DQVRNSYQLCNCNGLTRVMYANGGISFHSPEHVLGTTITIGRC	1765
Db	9	sslpssssfallstentqtrvrvnpdglrvtfasgmeiglssephilagavnpnlkgc	68
QY	1766	NISLPMENGLNSIEWRLRKEQIKGKVTIFGKRLRVHGNLLSIDYDRNIRTEKIYDDHRK	1825
Db	69	nislpghehnanlisvlxxgedgcaxnvfhlsfexhnnllisldfdhtrfgkydyddhrk	128
QY	1826	FTLRIYDQVGRPEFLWLPSSGLAAVNVSYFFNGRLAGLQRGAMSERDIDKQGRIVSRMF	1885
Db	129	ftlrllydqtrpdlwspvrynevnityspgslvtfgrgtwnekmeydqsflxspqlx	188
QY	1886	ADGKWSYSYLDKSMVLLQSQRYIFEYDSSDRLLAVTMPSPVARHSMSTHTSGYIRNI	1945
Db	189	lsilcysafsvfsqsmllhsqrryifeydqpdcillstmpsmvshsqtalsvgyyyni	248
QY	1946	YNPESNASVIFDSDGGRILKTSPLGTRGVQVFKYKGLSKLSEIVYDSTAVTFGYDETT	2005
Db	249	ytppdsstsfqdyrsrdgriqltlnhgrrrylykqkarlseviydtqtvtlycess	308
QY	2006	GVLMKVNLSQSGFSCTIRYKKGIPGVKQIYRSEBGMVNAERDYYTHDNSFRIASIKPV	2065
Db	309	gdlssdstliaxlltvtvlpagpligrqifrfseglvnmardfysy--nnfrvtsmqav	366

```
Qy 2066 ISETPLPVDLYRYDEISGVKVEHFGKGVYIYDINOITAVMTLSKHFDTHGRKEVOYE 2125
Db 367 inetplpdlryydvsgtrteqgkfsyindlqvtvttmkhtkifsangqvievqe 426
Qy 2126 MFRSLMYMTVOYDSMGRIKREKLKGPYANTTKYTDYDGGQLOLSVAVNDRPTWRYSY 2185
Db 427 ilkaiaymtlyqndvgrmvlcdlrvgvdantrfyeydagqlqtvsndkqwrtsy 486
Qy 2186 DLNGLHLLNPGNSVRLMPLRYDLRLDRITRLGQVQYIDDDGGLCCQSGSDIFEYNSKGLL 2245
Db 487 dlngnlnllshgksarltprydlrldrtrigeiqymdedgflrqgndlfeynsgll 546
Qy 2246 TRAYNKASGWSVOYRYDGVGRASYKYNLGHLLQYFYSDLNPRTHVYHNSSETISL 2305
Db 547 qkaynkasgwtvyydgdlgrvaskslgqlgfayadlcnprlvthlyntssetsl 606
Qy 2306 YYDLQGHFLFANESSGSEYVVASDNTGTPPLAVFSLINGLMIKLOLYTAYGEIYYDSNDFQ 2365
Db 607 yydlqghflfamelssgeeyvacdntgtplavfssrgqvikeillytpygdlyhdytpdq 666
Qy 2366 MVIGFHGLYDPLKLVHFTQRDYDLVLAGRWTSPTYTMKNVKGEPAPFNLYMFKSNPL 2425
Db 667 vilghaglydflcklvhlgqrdyvagrwttnphhiwkqlnllpkpfnlstklkygi 726
Qy 2426 SSELDLKNYVDKSWLVMFQSLNSIIPGPRAKMYFVPPPYELSESQ-----ASENGQL 2481
Db 727 fhflilcltdrswliefgqlhnlvpgfpkpelenspsixqmsnsmhlhllcaslxt 786
Qy 2482 ITGVQQTTERHNOAFMALEGQVITKHLHASTREKAGHW--PATTPPIIGKIMEAIREGR 2539
Db 787 ilgicelqklrnfisldqlpmcpryndgrcleggkqprfaavpsvfgkikfaikdgi 846
Qy 2540 VTTGVSSIASDSRKVASVLNNAYLLOKMHYSIEGKDTHYFVKIGSADGDLVTLGTIGR 2599
Db 847 vtadilgvanedsrrlaailnnahylenlhtiegrdthyfikigsleedvilgntggr 906
Qy 2600 KVLSEGVNVTYQPTLLVNGTRFRFTNIEFOYSTILLSIRYGLTPDLDEBKARVLQAR 2659
Db 907 rilengvntvsgntsvngtrrrfadiqlqhgalcfnairyg---ttveeeknhvleiar 963
Qy 2660 ORALCTAWAKQOKARDGREGSLRWTEGEKOLLSTGRVQGYGVYVLPVBO 2711
Db 964 qravaqatkeqrllqeeegirawtegekekqlstgrvgqdydgvfslsveq 1015

RESULT 8
B92780
AAB92780 standard; Protein; 625 AA.
AC AAB92780;
XX
XX
XX 26-JUN-2001 (first entry)
XX Human protein sequence SEQ ID NO:11266.
XX
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX Homo sapiens.
XX EP1074617-A2.
XX
XX 07-FEB-2001.
XX
XX 28-JUL-2000; 2000EP-0116126.
XX
XX 29-JUL-1999; 99JP-0248036.
XX 27-AUG-1999; 99JP-0300253.
XX 11-JAN-2000; 2000JP-0118776.
XX 02-MAY-2000; 2000JP-0183767.
XX 09-JUN-2000; 2000JP-0241899.
XX
XX (HELI-) HELIX RES INST.
XX
```

```
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI; 2001-318749/34.
DR
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602
XX full-length cDNAs defined in the specification, and for the detection
XX and/or diagnosis of the abnormality of the proteins encoded by the
XX full-length cDNAs -
PS Claim 8; SEQ ID 11266; 2537pp + CD ROM; English.
XX
XX The present invention describes primer sets for synthesizing 5602
XX full-length cDNAs defined in the specification. Where a primer set
XX comprises: (a) an oligo-dT primer and an oligonucleotide complementary
XX to the complementary strand of a polynucleotide which comprises one of
XX the 5602 nucleotide sequences defined in the specification, where the
XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX of an oligonucleotide comprising a sequence complementary to the
XX complementary strand of a polynucleotide which comprises a 5'-end
XX sequence and an oligonucleotide comprising a sequence complementary to a
XX polynucleotide which comprises a 3'-end sequence, where the
XX oligonucleotide comprises at least 15 nucleotides and the combination of
XX the 5'-end sequence/3'-end sequence is selected from those defined in
XX the specification. The primer sets can be used in antisense therapy and
XX in gene therapy. The primers are useful for synthesizing polynucleotides,
XX particularly full-length cDNAs. The primers are also useful for the
XX detection and/or diagnosis of the abnormality of the proteins encoded by
XX the full-length cDNAs. The primers allow obtaining of the full-length
XX cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
XX AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
XX AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
XX represent oligonucleotides, all of which are used in the exemplification
XX of the present invention.
XX Sequence 625 AA;
```

```
Query Match 16.0%; Score 2335; DB 22; Length 625;
Best Local Similarity 67.8%; Pred. No. 6.8e-122;
Matches 431; Conservative 91; Mismatches 94; Indels 20; Gaps 4;

Qy 2107 MTLSEKHDTHGRKEVOYEMPRSLMYMTVOYDSMGRIKREKLKGPYANTTKYTDYDQ 2166
Db 1 mtytkhfdahgrikeigveifrsimwyitqdmgrvtrkrelkigpantkyayevd 60
Qy 2167 DGOLOSVAVNDRPTWRYSYDLNGLHLLNPGNSVRLMPLRYDLRLDRITRLGQVQYIDDD 2226
Db 61 dggllgtvylneklmryndyngnlhllnpsnsarltprydlrldrtrigdvqyrlde 120
Qy 2227 GYLQCRGSDIFEYNSKGLLTRYANKASGWSVOYRYDGVGRASYKYNLGHLLQYFYSDLH 2286
Db 121 gflrqgteifeyskglltrvyksgwtviyrdglgrvssaktslgqlhffayadt 180
Qy 2287 NPTRITHVYHNSSETISLYYDLOGLHFLFANESSGSEYVVASDNTGTPPLAVFSLINGLMIK 2346
Db 181 yprtrithvynhsseitslyydlqghlfaelssgdefyasdtgtplavfssnglmk 240
Qy 2347 OLOQYTAIGEYIYDSNPQFQWIGFHHGLYDPLTKLVHFTQRDYDLVLAGRWTSPTYTMKN 2406
Db 241 qiytaygeiyfdsnldfqlvigfhgglydpltklihfgerdydlagrtwtptlelwr 300
Qy 2407 VGKEPAPFNLYMFKSNPLSSELDLKNYVTDVKSWLVMFQSLNSIIPGFRKMYFVPP 2466
Db 301 igkdpapfnlymfrnnpaskihdvkytdvnsvlvtfgfhlhnaipgfpvpkfdlcep 360
Qy 2467 PYELSESQASENGQLITGVQQTTERHNOAFMAL-----EGQVITKHLHASTREKAGH---- 2518
Db 361 syelvksgqwdidppifgvqqvargakafislgkmaevqv-----strragagqsw 412
Qy 2519 -WFATPTPIIGKIMEAIREGRVTGVSSIASDSRKVASVLNNAYLDDKMHYSIEGKDT 2577
Db 413 lwfatvksligkvmlavsggrvgtvnnianedcikaavlnnafylenlhtiekgkdt 472
```

QY 2578 HYFVKTSADGDLVLTGTTIGRKVLSEGVNVTVSQPTLLVNGTRTRFTNIEFYSTLLS 2637
Db 473 hyfikttesdgtlrlesgrkalenginvtsqstvtvngtrrrfadvemqfgalalh 532
QY 2638 IRYGLTPDILDEKARVLDOARALGTAWAKEQKARGREGSRUWTEGEKOQLLSTGR 2697
Db 533 vrygm---tldeekarileqarqalarawareqqrdrdgegarlwtegekrqlisagk 589
QY 2698 VQGYEGYVLPVQYOEYPELADSSNQFLRQNMGR 2733
Db 590 vqgygyvylsvsqypeladsannqlfrqselgrf 625
RESULT 9
AAU00392
ID AAU00392 standard; Protein; 429 AA.
XX
AC AAU00392;
XX
DT 04-JUL-2001 (first entry)
Human secreted protein, POLY4.
KW Human secreted protein; therapeutic; diagnostic; human; cancer.
OS Homo sapiens.
XX
PN WO200119856-A2.
XX
PD 22-MAR-2001.
XX
PF 13-SEP-2000; 2000WO-US25106.
XX
PR 13-SEP-1999; 99US-0153629.
PR 16-SEP-1999; 99US-0154520.
PR 20-SEP-1999; 99US-0154762.
PR 13-OCT-1999; 99US-0159231.
PR 12-SEP-2000; 2000US-0659634.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Shinkets RA, Fernandes E, Herrmann JL, Liu X, Yang M, Boldog FL;
XX
DR WPI; 2001-244781/25.
DR N-PSDB; AAS01213.
XX
PT New POLYX polypeptide useful for treating or preventing a POLYX
PT associated disorder, e.g. cancer
XX
YV Claim 9; Page 11-13; 152pp; English.
CC The sequence represents the amino acid sequence of human secreted
CC protein, POLY4. POLYX nucleic acids, polypeptides and antibodies to POLYX
CC can be used for treating or preventing a POLYX associated disorder in a
CC subject, preferably a human. These can be used in the manufacture of a
CC medicament for treating a syndrome associated with a human disease
CC selected from a POLYX-associated disorder, where the therapeutic is a
CC POLYX polypeptide, a POLYX nucleotide or a POLYX antibody. They may also
CC be used to screen for a modulator of activity, or latency, or
CC predisposition to a POLYX associated disorder, e.g. cancer.
XX
SQ Sequence 429 AA;
Query Match 14.6%; Score 2122.5; DB 22; Length 429;
Best Local Similarity 98.8%; Pred.No. 2.8e-110;
Matches 398; Conservative 0; Mismatches 4; Indels 1; Gaps 1;
QY 162 KSDDENGRPIPTSPFLPSAQLPSSHNPPVSCQMPLLDSNTSHQIMDTPDEEFSN 221
Db 10 kpsaeagrpiptspallpsaqlpsshnppvscqmpildntshqimdtndpdeefspn 69

QY 222 SYLLRACSGPQOASSGPPNHHHSQSLRPLPPPHNHTLSHHSSANSNRSLNRRSQ 281
Db 70 syllracsgpqqassgppnhsqslrplppphnhtlshhssanslnrsltnrrsq 129
QY 282 IHAPAPAPNDLATTPEVOLQDSWVLSNVPLETRHFLFKTSSGTPPLFSSSSPGYPLTS 341
Db 130 ihapapapndlattpesvqlqdswnlnsnvpletrhflfktsagstplfssspgyplts 189
QY 342 GTVYTPPPRLPRNTFSRKAFKLKPKSKYCSWKCAALSAAIAALLAILLAYFIVPWSLK 401
Db 190 gtvytppprllprntfsrkafklkpkskycwkcaalsaaiaaallaillayfivpwslk 249
QY 402 NSSIDSGEAEVGRRTVQEVPPGVFWRMSQIHSQOQFLKFNISLGKDALFGVYIRRLGPPS 461
Db 250 nssidsgaeavgrvrtvqevppgvfwrshsqpqflknislgkdalfgvyirrlgpps 309
QY 462 HAQYDFMERLDGKESVSVESPRERSIQTLVQNEAVFYQYLDVGLWHLAFYNDGKDKEM 521
Db 310 haqydfmerldgkesvsvesprersiqtlvqneavfyqyldvglwhlafyndgdkem 369
QY 522 VSFNTVVL-DSVQDCPRNCHNGECVSGVCHGCFPGFLGADCAK 563
Db 370 vsfntvvladvqdcprnchngecvsgvchcfcpgflgadca 412
RESULT 10
ABG07028
ID ABG07028 standard; Protein; 746 AA.
XX
AC ABG07028;
XX
DT 13-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #7019.
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR N-PSDB; AAS71215.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX
PS Claim 20; SEQ ID No 37387; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical

DR N-PSDB; AAS14084.
XX Novel polypeptides designated as FCTR polypeptides, useful in
PT detection, prevention and treatment of a broad range of pathological
PT states
XX
PS Claim 1; Page 33; 215pp; English.
XX
XX The invention relates to human FCTR polypeptides, FCTR1-FCTR7, and the
CC nucleic acids encoding them. These sequences are useful for the treatment
CC or prevention of numerous disorders including myelogenous leukemia,
CC carcinomas, melanomas, gliomas, astrocytomas, congenital neonatal
CC autoimmune thrombocytopenia, neurological disorders, neurodegenerative
CC disorders, nerve trauma, familial myelodysplastic syndrome,
CC Charcot-Marie-Tooth neuropathy, demyelinating Gardner syndrome, familial
CC myelodysplastic syndrome, mental health conditions, immunological
CC disorders, allergy and infection, bronchial asthma, Avellino type
CC eosinophilia, lung diseases, reproductive disorders, infertility, male
CC and female reproductive disorders, deafness, glycoprotein Ia deficiency,
CC desmold disease, turtot syndrome, liver cirrhosis, hepatitis C, gastric
CC disorders, pancreatic diseases such as diabetes, Schistosoma mansoni
CC infection Spino cerebellar ataxia, Plasmodium falciparum parasitaemia,
CC Corneal dystrophy-Greenow type I, Corneal dystrophy-lattice type I and
CC Reis-Bucklers corneal dystrophy. This sequence represents FCTR3a, a
XX neurestin-like protein.
XX
SQ Sequence 381 AA;

Query Match 13.2%; Score 1931; DB 22; Length 381;
Best Local Similarity 98.9%; Pred. No. 1.1e-99;
Matches 365; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 162 KSDDENGRPIPTTSSPLLPSAQLPSSHNPPVSCOMPLDLSNTSHQIMDNPDEEFSN 221
Db 10 kpsaeagrpipttsspsllpsaqilpsshnpvpsvcqmplldntshqimdnpeefspn 69

Qy 222 SYLLRACSGPOQASSSGPPNHHQSQTLPPLPPPHNHTLSHHSSANSLSNLSNRRSQ 281
Db 70 syllracsgpqgassgppnhhsqstlrpppphnhhtlshhssanslnrnslnrrsq 129

Qy 282 IHAPAPNDLATTPEVSQLODSWVLNSVPLETRHFLFKTSSGSTPLFSSSSGYPYPLTS 341
Db 130 ihapapndlattpesvqlqdswwlnsnvpletrhflfktsstplfssssgpyyplts 189

Qy 342 GTVYTPPRLPLRNTFSRKAFKLLKPKSKYCSWKCAALSAIAALLAILLAYFIVPWSLK 401
Db 190 gtvtytpprllprntfsrkafkllkpkskycskwcaalsaiaaallaillayfivpwslk 249

Qy 402 NSSIDSGAEVGRRTQEVPPGVFWRSQIHISQPOFLKFNISLCKDALFGVYIRGLPPS 461
250 nssidsgaeagrvtqevppgvfwrsgqihisqpfkfnislgkdalfgvyirglbps 309

Qy 462 HAQYDFMERLDGKKWSVVEPRRRSIQTLVONAEVVOYLDVGLWHLAFVNDGKKDEM 521
Db 310 haqydfmerldgkksvvesprrrsiqtlvqneavfvqyldvglwhlafyndgkdem 369

Qy 522 VSFNTVWLD 530
Db 370 vsfntvwd 378

RESULT 12
AAB61141
ID AAB61141 standard; Protein; 381 AA.
XX
AC AAB61141;
XX
DT 30-MAR-2001 (first entry)
XX
DE Human NOV11 protein.
XX
KW Human; NOVX; antinflammatory; cytostatic; neuroprotective;

KW cerebroprotective; immunomodulator; vulnerary; vasotropic; gene therapy;
KW hyperplasia; tumour; restenosis; psoriasis; Dupuytren's contracture;
KW diabetes; rheumatoid arthritis; cerebral oedema; Alzheimer's disease.
XX
OS Homo sapiens.
XX
PN WO200075321-A2.
XX
XX 14-DEC-2000.
XX
XX 01-JUN-2000; 2000WO-US15303.
XX
XX 03-JUN-1999; 99US-0137322.
XX
XX 16-MAR-2000; 2000US-0189810.
XX
XX 22-MAR-2000; 2000US-0191158.
XX
XX 30-MAR-2000; 2000US-0193086.
XX
XX 31-MAY-2000; 2000US-0137322.
XX
XX (CURA-) CURAGEN CORP.
XX
XX Shumkets RA, Fernandes E, Herrman J, Vernet C;
PI
XX
XX WPI: 2001-102403/11.
DR N-PSDB; AAF27859.
DR
XX
XX New NOVX polypeptides and polynucleotides, useful in gene therapy, as a
PT diagnostic marker, protein therapeutic, antibody or small molecule drug
PT target for treating immune, proliferative and metabolic diseases and
PT wound healing
XX
XX Claim 1; Page 42-44; 194pp; English.
XX
XX The present sequence is a new isolated polypeptide (NOVX). The NOVX
CC polypeptides, NOVX nucleic acids, and anti-NOVX antibodies are useful for
CC treating or preventing NOVX-associated disorders. They are also useful
CC for determining the presence of or a predisposition to a disease
CC associated with altered levels of the NOVX polypeptide or nucleic acid.
CC These NOVX-associated disorders include hyperplasias, tumours,
CC restenosis, psoriasis, Dupuytren's contracture, diabetic complications,
CC rheumatoid arthritis, cerebral lesions, diabetic neuropathies, cerebral
CC oedema, senile dementia or Alzheimer's disease. The NOVX polynucleotides
CC are especially useful in gene therapy. Specifically, NOVX is useful as
CC a diagnostic marker or prognostic marker, protein therapeutic and
CC antibody target or small molecule drug target to treat disorders in the
CC immune response pathway, thyroid and metabolic diseases, bone metabolic
CC disorders, diseases of the pancreas (e.g. diabetes or digestive
CC disorders), proliferative diseases, or tissue regeneration and
CC development (e.g. wound healing or treatment of burns).
XX
SQ Sequence 381 AA;

Query Match 13.2%; Score 1931; DB 22; Length 381;
Best Local Similarity 98.9%; Pred. No. 1.1e-99;
Matches 365; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 162 KSDDENGRPIPTTSSPLLPSAQLPSSHNPPVSCOMPLDLSNTSHQIMDNPDEEFSN 221
Db 10 kpsaeagrpipttsspsllpsaqilpsshnpvpsvcqmplldntshqimdnpeefspn 69

Qy 222 SYLLRACSGPOQASSSGPPNHHQSQTLPPLPPPHNHTLSHHSSANSLSNLSNRRSQ 281
Db 70 syllracsgpqgassgppnhhsqstlrpppphnhhtlshhssanslnrnslnrrsq 129

Qy 282 IHAPAPNDLATTPEVSQLODSWVLNSVPLETRHFLFKTSSGSTPLFSSSSGYPYPLTS 341
Db 130 ihapapndlattpesvqlqdswwlnsnvpletrhflfktsstplfssssgpyyplts 189

Qy 342 GTVYTPPRLPLRNTFSRKAFKLLKPKSKYCSWKCAALSAIAALLAILLAYFIVPWSLK 401
Db 190 gtvtytpprllprntfsrkafkllkpkskycskwcaalsaiaaallaillayfivpwslk 249

Qy 402 NSSIDSGAEVGRRTQEVPPGVFWRSQIHISQPOFLKFNISLCKDALFGVYIRGLPPS 461

|||||
Db 250 nssidsgeaevrrvtqevppgvtrsqihisqpfikfnislgkdaifgvvirglpps 309
Qy 462 HAQYDFMERLDGKEKSWVESPRRRSITQTLVQNEAVFOYLDVGLWHLAFYNDGKDEM 521
Db 310 haqydfmerldgkekswvesprrrsittqlvqneavfgyldvglwhlafyndgkdem 369
Qy 522 VSFNTVVLD 530
Db 370 vsfntvvld 378
|||||
RESULT 13
ABG04674
ID ABG04674 standard; Protein; 402 AA.
XX
AC ABG04674;
13-FEB-2002 (first entry)
Novel human diagnostic protein #4665.
DE Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.
KW
XX
OS Homo sapiens.
PN WO200175067-A2.
PD 11-OCT-2001.
PF 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
PI Drmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
DR N-PSDB; AAS68861.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
diagnostics, forensics, gene mapping, identification of mutations
responsible for genetic disorders or other traits and to assess
biodiversity
Claim 20; SEQ ID NO 35033; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
polypeptide (II) sequences. (I) is useful as hybridisation probes,
polymerase chain reaction (PCR) primers, oligomers, and for chromosome
mapping, and in recombinant production of (II). The
polynucleotides are also used in diagnostics as expressed sequence tags
for identifying expressed genes. (I) is useful in gene therapy techniques
to restore normal activity of (II) or to treat disease states involving
(II). (II) is useful for generating antibodies against it, detecting or
quantitating a polypeptide in tissue, as molecular weight markers and as
a food supplement. (II) and its binding partners are useful in medical
imaging of sites expressing (II). (I) and (II) are useful for treating
disorders involving aberrant protein expression or biological activity.
The polypeptide and polynucleotide sequences have applications in
diagnostics, forensics, gene mapping, identification of mutations
responsible for genetic disorders or other traits to assess biodiversity
and to produce other types of data and products dependent on DNA and
amino acid sequences. ABG0010-ABG30377 represent novel human
diagnostic amino acid sequences of the invention.
Note: The sequence data for this patent did not appear in the printed
specification, but was obtained in electronic format directly from WIPO
at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 402 AA;

Query Match 11.7%; Score 1708.5; DB 22; Length 402;
Best Local Similarity 92.7%; Pred No 3.1e-87;
Matches 331; Conservative 4; Mismatches 9; Indels 13; Gaps 2;
Qy 1899 SMVLLQSQRQYIFEXDSSDLLAVTMPVSVARHSMSTHTSIGYIRNIYNPPESNASVFD 1958
Db 27 smvllqsqrqyifeydssdrllavtmpsvvarhsmsthtsigyirniynppesnasvfd 86
Qy 1959 YSDGGRILKTSFLGTGRQVFKYKGLSKLSLSEIYVDSTAVTFGYDETGGVLMKNLQSGGF 2018
Db 87 ysdggrilksflgtgrqvfykglkslslseiyydstavtfgydettgvikmvlqsggf 146
Qy 2019 SCTIRYRKIGPLVDKQIYRFSEEGMVARFDYTVHDSFRIASIKPVISETPLPVDLYRY 2078
Db 147 sctiryrkigplvdqklyrfseegmvarfodytyhdnsfrlasikpvisetpplvdlyry 206
Qy 2079 DEISGKVEHFGKFGVIYIDINQIITAVMTLSKHFDHGRKEV-QYEMFRLMYWMTVQ 2137
Db 207 deisgkvehf-----fittaemtlskhfdhgrikevlayemfrslmywmtvq 254
Qy 2138 YDSMGRIKRELKLGVPYANTTKYDYDGGQLOSVAVNDPRTWRYSYDLNGLHLLNPG 2197
Db 255 ydsmgrikrclkigpyanttktydydggqlqsvacndrptwrysyelngnlhllnpg 314
Qy 2198 NSVRLMPLRYDLRDRITRLGSDVQYKIDDDGYLCQSGSDIFEYNSKGLLTRYNKAASG 2254
Db 315 nsvrlmplrydlrdritrlgldvqykiddgylcqsgsdifeynfgaptktsqsg 371
RESULT 14
ABG04673
ID ABG04673 standard; Protein; 242 AA.
XX
AC ABG04673;
XX
DT 13-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #4664.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
DR N-PSDB; AAS68860.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
diagnostics, forensics, gene mapping, identification of mutations
responsible for genetic disorders or other traits and to assess
biodiversity
Claim 20; SEQ ID NO 35032; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
polypeptide (II) sequences. (I) is useful as hybridisation probes,
polymerase chain reaction (PCR) primers, oligomers, and for chromosome
mapping, and in recombinant production of (II). The
polynucleotides are also used in diagnostics as expressed sequence tags
for identifying expressed genes. (I) is useful in gene therapy techniques
to restore normal activity of (II) or to treat disease states involving
(II). (II) is useful for generating antibodies against it, detecting or
quantitating a polypeptide in tissue, as molecular weight markers and as
a food supplement. (II) and its binding partners are useful in medical
imaging of sites expressing (II). (I) and (II) are useful for treating
disorders involving aberrant protein expression or biological activity.
The polypeptide and polynucleotide sequences have applications in
diagnostics, forensics, gene mapping, identification of mutations
responsible for genetic disorders or other traits to assess biodiversity
and to produce other types of data and products dependent on DNA and
amino acid sequences. ABG0010-ABG30377 represent novel human
diagnostic amino acid sequences of the invention.
Note: The sequence data for this patent did not appear in the printed
specification, but was obtained in electronic format directly from WIPO
at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 402 AA;

CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy involving
CC to restore normal activity of (II) or to treat disease states including
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published-pct-sequences.

	Query Match	8.6%	Score 1250;	DB 22;	Length 242;	
	Best Local Similarity	97.5%	Pred. No. 5.7e-62;			
	Mismatches 231; Conservative	1;	Mismatches 5;	Indels 0;	Gaps 0;	
Qy	2177 DRPTWRYSYDLGNLHLLNPGNSVRLMPLRVYDLDRITRGLGDVOVKIDDDGYLCORGSDI	2236				
Dd						
	5 drptwrysydingnl lpgnsvrmpirvydlrdrtrrlgavqykldddgylcgrgsdi	64				
Qy	2237 FEYNKSGLLTRAYNKASGWSQYYRDYGVRASYKTNLGHLLQYPFYSDLHPNTRITHVYN	2296				
Dd	65 feynskgl traynkasgwsqyyrdygvrasyktnlgllqlfyfysdlhnptrithvyn	124				
Qy	2297 HSNEITSLSYVDLOGLHFLAMESSGEYYVASDNTGTPLAVFSINGLMATKOLOYATYAGEI	2356				
Dd	125 hsneiitlslyyldggghlfamessgeeyvasdncgtplavfsinglmklqycaygei	184				
Qy	2357 YYDSNPDPQWVGIFHGGLYDLTLKLVLHFTRDYDVLAGRWTSPDYTMKNKGKEPAP	2413				
Dd	185 yydsnppdqwmvgifhggllydp tklvhtfqrdydvlagrtwspdytmknkvaggaqp	241				

RESULT 15
ABG03234
ID ABG03234 standard; Protein: 865 AA.
XX
AC ABG03234;
XX
DT 13-FEB-2002 (first entry)
XX
Novel human diagnostic protein #3225.

KW	Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW	food supplement; medical imaging; diagnostic; genetic disorder.
XX	
XX	Homo sapiens.
XX	
PN	WO200175067-A2.
XX	
PD	11-OCT-2001.
XX	
XX	30-MAR-2001; 2001WO-US08631.
XX	
PR	31-MAR-2000; 2000US-0540217.
PR	23-AUG-2000; 2000US-0649167.
XX	
XX	(HYSE-) HYSEQ INC.
PA	
XX	
PI	Dmanac RT, Liu C, Tang YT;
XX	
XX	WPI; 2001-639362/73.
DR	N-PSDB; AAS67421.
DR	
XX	

PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 20; SEQ ID No 33593; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at [ftp.wipo.int/pub/published_pct_sequences](http://wipo.int/pub/published_pct_sequences).
XX
XX Sequence 865 AA;

Query Match	8.3%	Score 1205:	DB 22;	Length 865;
Best Local Similarity	51.1%;	Pred.	No. 1.1e-58;	
Matches	282;	Conservative	24;	Mismatches 112; Indels 134; Gaps 171

QY	236	SSGPPNHHSOSTLRPPPLPHHNTLSHHHSANSLNRRSLNRRSQIHAPAPANDLATT	295
DB	158	sstapspbtittcrptcttp-----sanslnrrslnrrsqihnapapandlact	207
QY	296	PESVOLQDSWLVNSVPLETRHRFLFKTSSSGSTPLFSSSSPGYPLTSGVYTPPPRLPRN	355
DB	208	pessqlqdswnsvnpvletrhfiktsstgplfsssspgypvtsgtvytppprllprn	267
QY	356	TFSRKAFKLKPKSKYCSWKCAALSAIAAALLAILLAYFVPWSLKNSSIDSGEAE----	411
DB	268	tfsrkafklkpkskycwkcaalsaiagllaillayfiagpqmkrkaasqvdkrlr	327
QY	412	----VGRRTVTQEVPPGGVFWRFSQTHISQPQLKFNISLGKDALFGVYTRRGLP-----	459
DB	328	eillfvgnstvspfdd--wrgtv-----qemmgkckscfmferhmtlwtagl	375
QY	460	-----PSH-----AQDFMERLDGKEKWSVESPR	484
DB	376	iscilqqkqaagertcprhkaskqqlrtikkivdatydfmerlgdkekswvespr	435
QY	485	ERRSIQTLVONEAVFYQVLGVGLMHLAFYNDGCKEMVSFNVTVLDSVDQCPRNCHNGE	544
DB	436	errsiqtlvneavfydvglvhlafyndgkdkemvsfnvtvlsvdqdcprnchngne	495
QY	545	CVSGVCHCFGFLGADCAKAACPVLCSSONGOYSIGTCOCYSGWKGAECDVPMNQCIDPSC	604
DB	496	cvgvghcfpfiflgadcakgissvt-----qygsefdkik---kkasc--arnslalpv	545
QY	605	GGHGCSIDGNCVCVSAGYKGEHCBEVDCLDPTCSSHGVCNGECLCSPGWGGLNCELAR--	662
DB	546	vtags-----rqtaqlp-----ggfatlatret	569
QY	663	----VQCPRDQCGHGTLYLPDRGLCS-CBPNMWGPCSV-----EVCSVDGCTH-GVCIGAAC	713
DB	570	pitksqpdmprqp--pdtglnpgrhrhlepaagemrrriqlidsllhpsscfsatl	626
QY	714	RCEGWTGAACD	725

Db 627 kgka-aaacd 637
:| : ||||

Search completed: September 18, 2002, 11:08:02
Job time: 307 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 18, 2002, 11:05:30 ; Search time 48.75 Seconds
(without alignments)
5386.916 Million cell updates/sec

Title: US-09-800-198-8
Perfect score: 14581
Sequence: 1 MDVKDRRHSRLTRGRCGEC.....ELADSSNIQFLRQNMGRK 2733
Scoring table: BLASUM62
Gapop 10.0 , Gapext 0.5

rched: 283138 seqs, 96089334 residues
Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	9522	65.3	2825	2 T14271	Doc4 protein, stre
2	4465	30.6	849	2 T46253	hypothetical prote
3	3916.5	26.9	2515	2 S47008	tenascin-like prot
4	3637.5	24.9	2406	2 A54148	odz protein - frul
5	2208.5	15.1	2531	2 T16743	hypothetical prote
6	1171	8.0	782	2 A61625	tenascin-like prot
7	726	5.0	1810	1 A32230	tenascin precursor
8	680	4.7	2201	2 A32160	tenascin-C - human
9	679.5	4.7	2019	1 JQ1322	tenascin precursor
10	656	4.5	1746	1 S19694	tenascin precursor
11	632.5	4.3	4006	2 T09070	probable tenascin
12	627.5	4.3	184	2 T12457	hypothetical prote
13	616	4.2	3566	1 A40701	tenascin-X precurs
14	604	4.1	647	2 A43902	tenascin - eastern
15	590	4.0	4135	2 T42629	tenascin-x - bovin
16	463.5	3.2	2531	2 A46019	Notch-1 protein
17	462.5	3.2	2531	2 S18188	notch protein homo
18	452.5	3.1	3191	2 T22945	hypothetical prote
19	450.5	3.1	2318	2 S45306	notch 3 protein -
20	439	3.0	2555	2 A40043	notch protein homo
21	438.5	3.0	2321	2 S78549	notch3 protein - h
22	437.5	3.0	1064	2 A40136	fibropellin Ia - s
23	437	3.0	1620	2 T27283	hypothetical prote
24	434.5	3.0	1220	2 A56136	jagged protein pre
25	429.5	2.9	2471	2 A49128	cell-fate determin
26	428.5	2.9	2524	2 A35844	Notch protein - Af
27	421.5	2.9	1203	2 A49175	Notch B protein -
28	421	2.9	1356	2 A45445	Janusin precursor,
29	419.5	2.9	1574	2 T13954	MEGF6 protein - ra

ALIGNMENTS

RESULT 1

T14271

Doc4 protein, stress-induced - mouse

N:Alternate names: odz protein homolog

C:Species: Mus musculus (house mouse)

C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999

C:Accession: T14271

R:Wang, X.Z.; Kuroda, M.; Sok, J.; Batchvarova, N.; Kimmel, R.; Chung, P.; Zinszner,

EMBO J. 17, 3619-3630, 1998

A:Title: Identification of novel stress-induced genes downstream of chop.

A:Reference number: Z17951; MUID:98315054

A:Accession: T14271

A:Status: preliminary; translated from GR/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-2825 <MAN>

A:Cross-references: EMBL:AF059485; NID:g3170614; PID:g3170615; PIDN:AAC31807.1

C:Genetics:

A:Gene: Doc4

Query Match 65.3%; Score 9522; DB 2; Length 2825;
Best Local Similarity 61.6%; Pred. No. 0;
Matches 1763; Conservative 421; Mismatches 511; Indels 168; Gaps 28;

Qy 1 MDVKDRR-HRSLTRGRCGECRYTSSLDSDCRVPTQKSYSSSETLKAYDHDSDRMHYGN 59

Db 1 MDVKERPYSRLTRRR-DAERRYTSSSADSEEGKP-QKSYSSSETLKAYDQDARLAYGS 58

Qy 60 RVTDLIHREDEFFPRQGTNETLAELGICEPS-PHRSGYCDMGLHGQYSLGSDADSD 118

Db 59 RVKDMVPQEAEEFCRTGNTFLRELGLGEMTPPHGTLRYTDIGLPHCGYSGAGSSADADLE 118

Qy 119 TEGGMSPEHAIRLWGRIGKSRSSGLSSRENSALTITDSDNENKSDDENGRPIPTSSPS 178

Db 119 ADTVLSPEHPVRLWGRSTRSGRSCSSRANSNLTDTEHEN---TETGAPL-HCSSAS 174

Qy 179 LLPSAQPLSSHNPPVPS-CQPLLDSTNSHOIMTNDPEEPSNYSLLRACSGPQQQASS 237

Db 175 STPIEQSP 234

Qy 238 GPP----NHHSQSRLRPLPPPHNHTLSHHH-SSANSINLNRSLNRRSQIHAPAPAPND- 291

Db 235 DHPSSLQNHPLRLP--PPPLPHAPHTPNQHAASINLNRGNFTPRSN----PSPAPTDH 288

Qy 292 -----LATTPESVQLQDSVNLNSNPLETR----- 316

Db 289 SLSCPPAGSAQEPHQAODNVLNSKIPVETRNLGKOPFLGTWQDNLNLIEMDIFSASRRDG 348

Qy 317 -----HFLFKTSSGSTFLFSSSSPGYPLTSGTYVTPPRLPLPRNTFSRKAFLKPKSKYC 371

Db 349 AYSDGHEFFK-PGCTSPFLCTTSPGYPLTSTVYSPPPPLPRPLRSTFSRPAFNLAFLKPKSKYC 407

crumbs protein - f
transmembrane prot
notch protein - fr
restricitin precurs
notch homolog - se
Notch homolog prot
gip1 protein precu
neurogenic protein
neurogenic repetit
gene serrate prote
gene Delta protein
hypothetical prote
notch4 - mouse
Notch homolog Motc
adhesive plaque pr

Db 2550 TDAMEPSYELVHTQMKTOEMDNKSIILGVOCEVQKQKAFVTLERFDQLYGSTIISCOQA 2609
Qy 2511 SIREKAGHWFATTTPIIGKIMFAIKEGRVTTGVSSIASEDSRKVASVLLNAYYLDKWHY 2570
Db 2610 PETKK- - - - -PASSGIFGKGVKFKDGRVTTDIISVANEDGRRRAAIIINAHYLENHUF 2665
Qy 2571 SIEGKDTTHYFKVIGSADGLVTLTGTTIGRKVLESGVNVTVSQPTLLLVNGTRRRFTNIEFQ 2630
Db 2666 TIDGVDTHYFKVPGPSEGLAILGLSGGRRRTLENGVNVTVSQINTMLSGRRRTYDIQLQ 2725
Qy 2631 YSTLLSRYGLTPDTLDEEKARVLDQARQALGTAWAKEQKARDGREGSRRLTEGKQ 2690
Db 2726 YRALCLNRYG- - - - -TVDDEKVRVLELQARAVROAREOQRLREGEGELRAWTDGKQ 2782
Qy 2691 OLLSTGRVQGVYGVYVLPVEQYPELADSSNIQFLQONEMGR 2733
Db 2783 QVLTNTRVQGVYGVYVLPVEQYPELADSSNIQFLQONEMGR 2825

RESULT 2
T46253
hypothetical protein DKFp761F171.1 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000
C:Accession: T46253
R:Ottenwaeider, B.; Obermaier, B.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, January 2000
A:Reference number: 223031
A:Accession: T46253
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-849 <AAA>
A:Cross-references: EMBL:AL137500
A:Experimental source: adult amygdala; clone DKFp761F171
C:Genetics:
A:Note: DKFp761F171.1

Query Match 30.6%; Score 4465; DB 2; Length 849;
Best Local Similarity 100.0%; Pred. No. 7.4e-230;
Matches 849; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1885 FADGKWSYSLDKSWLLLSQROYIFEYDSSDRLLAVTMPVSARHSMSTHTSGYTRN 1944
Db 1 FADGKWSYSLDKSWLLLSQROYIFEYDSSDRLLAVTMPVSARHSMSTHTSGYTRN 60

Qy 1945 IYNPPESNASVIFDYSDGRILKTSFLGTGRQVYKYKLSLSEIVYDSTAVTFGYDET 2004
Db 61 IYNPPESNASVIFDYSDGRILKTSFLGTGRQVYKYKLSLSEIVYDSTAVTFGYDET 120

Qy 2005 TGVLMVNLQSGGFCTIRYRKIGPLVDKQIYRFSEEGMVNARFDYTYHDNSFRASIATKP 2064
Db 121 TGVLMVNLQSGGFCTIRYRKIGPLVDKQIYRFSEEGMVNARFDYTYHDNSFRASIATKP 180

Qy 2065 VISETPLVDLYRYDEISGVEHFGKGVYIYDINQIITAVMTLSKHPDTHGRKIEVOY 2124
Db 181 VISETPLVDLYRYDEISGVEHFGKGVYIYDINQIITAVMTLSKHPDTHGRKIEVOY 240

Qy 2125 EMFRSLMVMVTVOYDSMGRTIKRELKGLPYANTTKYTYDYGDOLOSVAVNDRPTWYS 2184
Db 241 EMFRSLMVMVTVOYDSMGRTIKRELKGLPYANTTKYTYDYGDOLOSVAVNDRPTWYS 300

Qy 2185 YDLNGLHLLNPGNSVRLMPLRYDLRDLRITRLGDYQYKIDDDGYLCQSGDIFEYNSKGL 2244
Db 301 YDLNGLHLLNPGNSVRLMPLRYDLRDLRITRLGDYQYKIDDDGYLCQSGDIFEYNSKGL 360

Qy 2245 LTRAYNKASGWSQVRYDGVGRRASYKTNLGHLLQYFYSDLNPNTRITHVYHNHNSITTS 2304
Db 361 LTRAYNKASGWSQVRYDGVGRRASYKTNLGHLLQYFYSDLNPNTRITHVYHNHNSITTS 420

Qy 2305 LYDLOGLHFLAMESSGEYVYASDNTGTPPLAVFSINGLMIKOLQYATGEIYYDSNPDF 2364
Db 421 LYDLOGLHFLAMESSGEYVYASDNTGTPPLAVFSINGLMIKOLQYATGEIYYDSNPDF 480

Qy 2365 QMWIGFHGGLYDPLTKLVHFTQRDYDVLAGRWTSPDYTMKNVKGEPAPFNLYMFKSNPP 2424
Db 481 QMWIGFHGGLYDPLTKLVHFTQRDYDVLAGRWTSPDYTMKNVKGEPAPFNLYMFKSNPP 540

Qy 2425 LSSELDLKNYVTDVKSWMVFGFQLSNIPGPPRAKMYFVPPPYELSSQASENGOLITG 2484
Db 541 LSSELDLKNYVTDVKSWMVFGFQLSNIPGPPRAKMYFVPPPYELSSQASENGOLITG 600

Qy 2485 VQOTTERHNQAFMALEGQVITKKLHASIREKAGHWFATTTPIIGKIMFAIKEGRVTTGV 2544
Db 601 VQOTTERHNQAFMALEGQVITKKLHASIREKAGHWFATTTPIIGKIMFAIKEGRVTTGV 660

Qy 2545 SSIASEDSRKVASVLLNAYYLDKMHYSIEGKDTHTYFVKIGSADGLVTLGTTIGRKVL 2604
Db 661 SSIASEDSRKVASVLLNAYYLDKMHYSIEGKDTHTYFVKIGSADGLVTLGTTIGRKVL 720

Qy 2605 GNVTVSQPTLLVNGTRRRFTNIEFOYSTLLLSIRYGLTPDTLDEEKARVLDQARQALG 2664
Db 721 GNVTVSQPTLLVNGTRRRFTNIEFOYSTLLLSIRYGLTPDTLDEEKARVLDQARQALG 780

Qy 2665 TAWAKEQKARDGREGSRRLWTEGKQQLLSTGRVOGYEGYVLPVEQYPELADSSNIQF 2724
Db 781 TAWAKEQKARDGREGSRRLWTEGKQQLLSTGRVOGYEGYVLPVEQYPELADSSNIQF 840

Qy 2725 LRQNEGMGR 2733
Db 841 LRQNEGMGR 849

RESULT 3
S47008
tenascin-like protein - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 27-Jan-1995 #sequence_revision 27-Jan-1995 #text_change 21-Jul-2000
C:Accession: S47008
R:Baumgartner, S.; Martin, D.; Hagios, C.; Chiquet-Ehrismann, R.
EMBO J. 13, 3728-3740, 1994
A:Title: Ten(m), a Drosophila gene related to tenascin, is a new pair-rule gene.
A:Reference number: S47008; MUID:94349920
A:Accession: S47008
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-2515 <BAU>
A:Cross-references: EMBL:X73154; NID:9510505; PIDN:CAA51678.1; PID:9510506
C:Genetics:
A:Gene: FlyBase:Ten-m
A:Cross-references: FlyBase:FBgn0004449
C:Superfamily: unassigned EGF-related proteins; EGF homology
F:298-324/Domain: EGF homology <EGF>
F:394-421/Domain: EGF homology <EGF>

Query Match 26.9%; Score 3916.5; DB 2; Length 2515;
Best Local Similarity 34.7%; Pred. No. 7.8e-200;
Matches 836; Conservative 475; Mismatches 922; Indels 177; Gaps 52;

Qy 408 GEAEGRVRTQVPPGVFWRSOIHIISQOFLKFNISLGDALFVYIRGLPPSHAQYDF 467
Db 177 GOITLQGLKTRIQYISYWNMOFYOSEPAYRFDITPRGASIGYGRNALPTHTQYHF 236

Qy 468 MERLDGKEKWSYVESPRRRRSIQTLLVQNEAVEVQYLDVGLMHLAFYNDGKD-KEMVSFT 526
Db 237 KEVLSS- - - - -PSASTRTARAHLSTIREV- - - - -TRYMEGHWFVSLYNDGDVQELTFYAA 289

Qy 527 VVLDVSQDCPRNCHNGECVSGVCHCFPGFLGADCAKACPVLCSGNGOYSGTKCQCYSG 586
Db 290 VAEDMTQNCPNGCSGNGOYSGVCHCFPGFLGADCAKACPVLCSGNGOYSGTKCQCYSG 349

Qy 587 WKGACEDVPMNOCIDPSGGHGCIDGNCVCSAGYKGEHCHEVDCLDPTCSHGGVGVNAGE 646
Db 350 WKGECESLRHDEADVDCSGHGHCVSGQCQMRGKYGKFCFEEVDCPHPNCSGHGFCADGT 409

QY 647 CLCSPHGGNLC---ELARVQCPCQCGHGTCTYLPDGTGLCSCDPNWMGPDGCSVEVCSVDCG 703
Db 410 CICKKMGKPCDCAIMQDALQCTCPDSCGHGTCTCTCEAKVSGDDCKSKELCDLDCG 469
QY 704 THGVCJGACRCREGWTGAACDORVCHPRCIEHGTCKDKGCECREGWNGBHCTIGRGTAG 763
Db 470 QHRCRGDADACDPENGGEYCNTRLCOVRCNEHQCKNGTCLCTVGNGBKHCTI----- 533
QY 764 TETDGPDLGNGRCTL-QGNSWCVCQGTGRGPCNVMAMETSCADKNKNEGDGDLCL 822
Db 524 ---EGCPNSCAGHCQCRVSGEGWECRCYEGWDPGDCGIALELNCGDSKDNKDLGVDC 580
QY 823 DPCCLOSAACONELLRCGRSDPLDIIOGQTDWPAV-KSFYDRIKLAGKDSHTIIPGEN 881
Db 581 DPCCASHVCKTSQLCVSAKPTDVLKQP--PAITASFERRMKFLIDESSLQNTAKLE 638
QY 882 PFNSLSVLIRGOVTTDGTPLGVNVSVFKYPKYGTITRODGTDLTANGCASLTLP 941
Db 639 TFNESRAVIRGRVTVSLGMLGVVRVSTTLL-FTLTRDDGNFDMVNGGAVTLQF 697
QY 942 ERAPFMSQERTVWLPNWSFYAMDTLVMKTEE-----NSIPSCDLSGFVRPDP 995
Db 698 GRAPFRPQSRIVQVPNVEVVIIDLVMNMSSEKGLAVTTTTCFAHDYDLMKPVVLASWK 757
QY 996 STEFSAAPGONPIVETQVLHBEIPLPGSNVKRLYLSSRTAGYKSLKITMTOSTVPLNL 1055
Db 758 HGOGACPDORSALAESQVLESQIPGTGLNLVYHSSRAAGYLSITIKQLTDPDVPTSL 817
QY 1056 IRVHLMAVBEHGFQKFSQFASPLNASTFIWDKTDAYGORVYGLSDAVSVGFYEYTCPSL 1115
Db 818 HLJHLRITIGILFERIFEADPGIKETAYAWNRLNIYRQYGYTTAVKVGYYTCTD- 876
QY 1116 ILWEKRTALQGFELDPNSLGNWSLKKHILNLYKSGILHKGTCGENOFLTQQAIIYSIMG 1175
Db 877 IYWDIQTTKLSGHDMSTSEVGGWNLTHRYNPHEGTLQKDGCSNLYLRNKPRIILLTGM 936
QY 1176 NGRRRSISPCSCNGLAEGNKLAPALAVGIDGLSVYGFNFYIRRRFPNVTSLLELRN 1235
Db 937 DGHORLEPCDCQATQKRLAPALAAAPDGLSVGFNFYIRRMTPDGSIRTVKVL-- 994
QY 1236 KEFKHNNPAHYLLAVDPVSGSLYVSDTNSRRIRYVKLSGTKDLAGNSEVAVAGTEOC 1295
Db 995 ----NATRVSYRHYMALSPDLGTLVSDPESHQIIRVROTDNYSQPELWAEVVGSGERC 1050
QY 1296 LPDEARCGDGGKAIDATLMSPRGIAVDKNGLMFYVDATMIRKVDQNGIISTLLGSN-DL 1354
Db 1051 LPGDEAHCGDGALEKAKLAYPKGIAISSDNILYFADGTNIRWVDRDGIYSTLIGNMHK 1110
QY 1355 TAVRPLSCDSSMDVAQVRLEWPTDLAVNPMDNSLYLVLENNVILRITENHOVSIAGRPMH 1414
Db 1111 SHWKPIPCBGTLEEMHLRWPTLAVSPMDNTLHIDDHMLRMTPDGRVRVISGRPLH 1170
QY 1415 CQVPGIDYLSKLAHSALESASAIASHGTGVLYIPETDEKKINRLROVTTNGEICLLAG 1474
Db 1171 CATASTAYD-TDLATHTLVMPQSIAGPLGELYVAESDSQRINRVITCDGRIAPFAG 1229
QY 1475 AASDCCKNDVNCYSGDDAYATDAILNPSPLAVAPDGTIYIADLGNIRAVSNKRP 1534
Db 1230 AESKCNCL-ERGCDCFEAEHYLATSAKFNTIAALAVTPDSSHVHIADQANYIRSVMSSTP 1288
QY 1535 VLNAFNOYEAASPGOEYLYFVNADGTHQVTSVLTGEYLYNFY---STDNDVTELIDNN 1591
Db 1289 EASPSREYIAPDMQOEYIFENRFGHVSTRNLTGETTVFTYVNTSNGKLSVTDDA 1348
QY 1592 GNSLKIRRDSSGPHRLLMPDQNIITLVCTNGSLGWSTQNLGLMTYDGTGLLATK 1651
Db 1349 GNVKVELLOYTSQVNSIENTKGKCLRWTRMKMLHELSTPDPNVNYEYHGTGLLRK 1408
QY 1652 SDETGWTFYDHDHGRNLNVRTPGTGVVTSLHREME-KSITIDI-ENSNRD-----DDVT 1704
Db 1409 LDSTGRSYVNYNDEFGRLTSAVTPTRGVRIELSPDLKVAQVKSNAQKEMSLITQAT 1468
QY 1705 VITNLSSVEASYTVVQDVRNSVQLNCTLRVMYANGMGSIFHSEPHVLAGTITPTIGR 1764

Db 1469 VIVRNAAASRTTVDMD-----GSTTITPMGHNLQMEVAPYTTILABQSPLLGE 1517
QY 1765 CNLSLPE-----NGLNSTEW--LRKE-----OIKG---KVTIFGRKLVRHGR 1803
Db 1518 ---SYPPAKQRTIEAGDLANREFWRYFVRROQPLAQKOSKGPVRPTEVGRKLVRNGD 1574
QY 1804 NLSIDYDRNRIRTEKIYDHRKFTLRIYDQVGRPFLWLPSSG-LAAVNVSYFNGRLAG 1862
Db 1575 NVLTLEYDRETQSVVMVMDKQELLNVYDRTSPISFRPQSGDYAVVDLEYDRFGRLV 1634
QY 1863 LQRCANSERTDIDQGRIVSRMFADKVMWSYSYLDK---SMVLLQSORQ---YIFEVDSSD 1918
Db 1635 WKGVLOEAYSFORNRLNEIKYGDGSTMVYAFKDFGSLPLKVTTPRRSDYLLQYDDAG 1694
QY 1919 RLLAVTMSVARHSMTHTSIGVIRNLYNPENASVIFDYSDDGRILKTSFLTGROVF 1978
Db 1695 ALOSLLTPRGHIIHAFSLOTLGFFKYQYYPINRHFELIYNDGQILAKIHPHQSKVA 1754
QY 1979 KYGKLSKSEIYVDSTAVTFYDGTGVLKVMNLSQSGFSCSTIRYR-KIGPLVDKQIYR 2037
Db 1755 FVHDTAGRLLETILAGLSSTHYTYDQTTSLVKSVEVOBPGPELRRERFKYHAGILKDEKLR 1814
QY 2038 FSEEGVNAFEDYTYHDNSFRIASIKPVISETPLVDLYRYDEISGKVEHFGFVYIID 2097
Db 1815 GSKNSLASARIKYAY-DGNARLSGIEAIDDKELPTTRYKYSONLGOLE-----VYOD 1866
QY 2098 I-----NQIITAYMTLSKHF-----DTHGRKEVOYEMERSLMYMTVOYDSMGRVIKR 2147
Db 1867 LKTRNAFRTVQDSAKQFFAIVDYDOHGRVKSVMNKNIDVFRLEDDYDLNRKKSQ 1926
QY 2148 ELKGPYANTKTYDYDGGQLQSVAVNDRPTWRYSYDLNGN-LHLLNPGNSVRLMPLR 2206
Db 1927 KTTFGRTAFDKINYADG---HVVEVLGNTNMKYLDFENGNTVGVVDOGEKFN---LG 1979
QY 2207 YDLDRITRLGDVOY-KIDDDGYLCORGSDIFEYNSKGLLTRYANKA--SGWSVQVRYDG 2263
Db 1980 YDIGDRVIVKGVDFEYNNYDARGFVVRKGEQKYRNNRGLIHSFERFERFSW---YYDD 2036
QY 2264 VGRASYKTNLGHLLQYFYSDLNHPTRITHVYNHNSSETLSYDYDQGHFLHFAESSGEE 2323
Db 2037 RSLVAVHNDKNGTQYYANPRTPHLVTHVHPFKISRTMKLFYDDRDMLIALEHD-QR 2095
QY 2324 YVVASONTGTPLAVFINGLMIKOLQYATGEYIYDSNPDQFQWVIFGHGLYDPLKLVH 2383
Db 2096 YVYATQNGSPLAFFQNGSIVKEMKRTPGRIKDTKPEFFVPIDFHGGLIDPHTKLVY 2155
QY 2384 FTORDYDLAGRWTSPTYTMKNVKGK---PAPFNLYMEKSNPNLSELDKNYVTDVKS 2440
Db 2156 TEORQYDPHVQWMTB---LWETLATEMSHTDVFYIRYHNNDPINPNKP-QNYMIDLDS 2311
QY 2441 WLVMFGOLSNITIPFPRAKMYEVPYPPYELSESQASENGOLITGVQQTTERHNAQFM--- 2497
Db 2212 WLQFCYDLNNMSSRYTKLAQYTPQASIKSNLAPDFG-VISGLECIKTESEKSDPD 2270
QY 2498 -----ALEQVITPKLHASIREKAGHWFATTTPIIOGKIMFAIKEGR---VVTGVSSIA 2548
Db 2271 FVPKPLLTKEPKNRNLPRVSYRRG-----VFGEVLLSRIGGRALVSVVDGNSV 2322
QY 2549 SEDSRKVASVNNAYVYLDKMHYSEICKDTHYFV-----KIGSADGDLVLTGTTIGRKVLE 2603
Db 2323 QD---VSSVFNNSYFELD-LHFSIHQDQVYFVKONVVKLRDDNEELRGGMFNTSTHE 2378
QY 2604 SGVNVTVSPTLLVNGTRRFRFTNIEFYSTLLSIRYGLTPTDPLEKARVLDQARQAL 2663
Db 2379 ISDHGCSAAKELRLHG-----PDADVIIKYGVDP---EOERHRLKHAHKRAV 2423
QY 2664 GTAWAKEQKARDGRGSRBLWTEGEKQQLLSTGRVOGYEGYVLPVQYPELADSSNTQ 2723
Db 2424 ERAWELEKOLVAAGFGRGDGTBEKEELVQHDGVDGWNIGDIHSHKYPQLADDPGNA 2483
QY 2724 ELRQNEWGR 2733

Db 2484 FORDAKRRR 2493

RESULT 4

A54148

od2 protein - fruit fly (*Drosophila* sp.)

C:Species: *Drosophila* sp.

C:Date: 02-Aug-1994 #sequence_revision 02-Aug-1994 #text_change 11-Jan-2000

C:Accession: A54148

R:Levine, A.; Bashan-Ahrend, A.; Budai-Hadrian, O.; Gartenberg, D.; Menasheerow, S.; Wide

Cell 77, 587-598, 1994

A:Title: odd Oz: a novel *Drosophila* pair rule gene.

A:Reference number: A54148; MUID:94243925

A:Accession: A54148

A:Status: preliminary; not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-2406 <LEV>

C:Genetics:

C:Gene: FlyBase:Ten-m

C:Cross-references: FlyBase:FBgn0004449

C:Superfamily: unassigned EGF-related proteins; EGF homology

F:514-540/Domain: EGF homology <EGF>

F:610-637/Domain: EGF homology <EGF1>

Query Match 24.9%; Score 3637.5; DB 2; Length 2406;

Best Local Similarity 32.9%; Pred. No. 5.3e-185;

Matches 817; Conservative 444; Mismatches 901; Indels 325; Gaps 59;

QY 165 DENGCRPTPTSSPLPSAQLPSHNPVPSVCOMPLLDSNT-SHQIMDTNPDEEFSNSY 223

Db 14 DVGGGTPGHAHPH-AQGRTLPLMSSHGRRPTDLGPHVHGSQTLOHQ-----NQ 60

QY 224 LLRACSGPOAAS-----SSGPP---NHHQSOT-----LRPPLPPPHN 258

Db 61 NLQAVQAASSHYDYEYQHHLAHPPTVNTAQRTHGROGFLLEGVTPATPDVPPRNP 120

QY 259 TLSHHSSANLNRNSTNRSSQIHAPAPA---PNDLATTPEVSQLODSWYLSN---VPL 313

Db 121 TMSR-----MONGRLTVNNPDADFPSCVLRTVPSGNVTPSGNINLNGKSPI 168

QY 314 ETRHFLFKTSSGTPLEFSSSPGYPLTSGVYTP--PPRL-----LPNRTSRKAFKLKP 367

Db 169 D-----FKSGSACSTPTKDTLKGYERTSGCMGPVLQPSRVNGLPAHHYSAPMFRKDL 223

QY 368 SKYCS---WKCAALSAIAAALLAILLAYF-IVPWS----- 399

Db 224 VARCSPWFSGISVLFVAFVWMLILTTTGVTKHNSPPCSVLVNGEASEVTAAKSTNT 283

QY 400 ----LNKSSI-----DSGEAE----- 411

Db 284 DLSKLHNSVRAKNGOGIGLAQSGGLGAAGVGGGSSAATVTTATNSGTAGQLQSTS 343

QY 412 -----VGRVTO 418

Db 344 ASAEATSAATSSQSLSLTSLSSLANANGARTPAKSFPPDGTFTGOITLGLQKLTK 403

QY 419 EVPPGVFWRSQIHISQPOFLKFNISLGLKDALFGVYIRRGLPSPSHAQYDFMERLDGKEKWS 478

Db 404 EIOPYSYNNQFYQSEPAYVKFDYTPRGASIGVYGRNALPTHQYHFEVLSC----- 458

QY 479 VVESPERRSTIQTULQNEAVFVQYLDVGLWLHLPYNDGKD-KEMVSFNTVVVDSVQDCPR 537

Db 459 FSASTRARAHLSTIREV--TRYMEPGHFWFSVLYNDGDVQELTFYAAVAEDMTQNCN 516

QY 538 NCHNGECVSVCHCFPGFLGACADCAKACPVLCSGNGQYSGKTCQYSGWKGAECVPMN 597

Db 517 GCSGQGLLGHCCNCFPGGDDCSVCPVLCSQHGLEYTNGECICNPWKMKECESLRHD 576

QY 598 QCIDPSCGGHSGCIDGNCVCSAGYKGBHCEBEVDCLDPTCSHGVVNGECLCSPGWSGLN 657

Db 577 ECEVADCSGHGHCVSGKQCQMRGKYGKFCBEVDCPHPCNSGHHGFCADGTCICKKGWKGPD 636

QY 658 C---ELARVQCPDQCSCGHGTYLPDTGLCSDPNMMGSDPCSVEVCSVDCTGHGVCIGACR 714

Db 637 CATMDQDALCQLPDCSGHGTFLDTQTCTCEAKWSGDCSKELCDLDCGQHGRCGDACA 696

QY 715 CEEGWTGAACDQQRVCHPRCIEHGTCTDGKCECEGNGEHCITIGRTAGTETDGCPLCN 774

Db 697 CDEWGEYCNTRLCDDVRCHNEHGCKNGTCLCVTGNNGKCTI-----EGCPNSCA 747

QY 775 GNGRCTL-GONSQCVCOTGRPGCNVAMETSCADNKDNEGLVCLDLPDCCLOASCO 833

Db 748 GHGCRVSGGWECEYEGWDPDCALELNCGDKNKDKGLVDCDEPECCASHVCK 807

QY 834 NSLLCRGRDPLDIIQOQTDWPAV-KSFYDRIKLLAGKDSHTIIPGENPNSLSVLIR 892

Db 808 TSOLCVSAPKPIDVLLRKQP--PAITASFFERMKFLIDESSLQNYAKLETFNESRAVIR 865

QY 893 GOVVTDTGTLVGVNVSFVYKPYGTYITRODGTFLDIANGASLTLHFERAPMSERT 952

Db 866 GRVYTSLGMGLVGRVSTTTLLLE-GFTLTRDDGDFLVMNGGAVTLQFGRAPERQPSRI 924

QY 953 VLPWNSFYAMDITLVKTEE-----NSIPSCDLSGVRPDPPIIISPLSTFFESAAPQN 1006

Db 925 VQVPWNEVVIIDLVMMSSEKGLAVTTHTCFADHYDLMKPVVLAWSKKHGFQACPDRS 984

QY 1007 PIVPETQVLHEIEELPGSNVLYLRSRTAGYKSLKLTMTQSTVPLNLRVHLMVAEG 1066

Db 985 AILAESQVIOESLQIPGTGLNVVHSSRAAGYLSTIKLQLTDPDPTSLHLHLIRITIEG 1044

QY 1067 HLQKFSQASPNLASTPIWKTDAYQORVYGLSDAVYSGVEYETCPSLILWEKRTALLO 1126

Db 1045 ILFERIFEADPGKRTFYAWNRNIYRQVYGVTTAVYKVGQYQYDCTD-IVMDIQTTKLS 1103

QY 1127 GFELDPNGLGWSLKDHHILNVKSGILHKGTTGENQFTQTPAITTSMGRRRSISCPSS 1186

Db 1104 GHDMSEVGGWNLDIHRYNFHEGILQKGDGSIYLRNKPRIILITMGDGHQRPCEPD 1163

QY 1187 CNGLAENKLLAPVALADIGSLYGVDFNYIRIFPSRVNTSLELRNKEFKHNSNPAH 1246

Db 1164 CDGQATKORLLAPVALAAPDGLFVGDFNYIRIMTDGSGTIVVKL-----NATRVSY 1217

QY 1247 KYVLAVDPVSGSLVSDTNSRRLVYRVKLSGTDKLAGNSEVAVAGTGEQCLPFDEARGDG 1306

Db 1218 RYHMAUSPLDGLTVSDPESHQIIRVTDNDYSQPELNWEAVVGGSERCLPDDARCGDG 1277

QY 1307 GKATDAPLMSPRGAVDKNGLMYFVDATMIRKVDONGIITLLGSN-DLTAVRPLSCDSS 1365

Db 1278 ALAKADAKLAPKGLAISDNILYFADCTNIRVMDRDCIVSTLIGNHHMKSHWKPIPEGT 1337

QY 1366 MDVAQVLEWPTDLAVNPMDSLVVLENNVILRTENHQVSIAGRPMHCOVPOIDYSL 1425

Db 1338 LKLEEMHLRWPTELAVSPMDNTLHIIDHMLRMTPDGRVVRISGRPLHCAATASTAYD-T 1396

QY 1426 KLATHSALESASATAISHTGVLYITETDEKKINRLROVTTNGEICLLAGASDCDCKNDV 1485

Db 1397 DLATHATLVMPQSTAFGLGELYVAESDSORINRVVIGTDGRIPAPFAGAESKNCNL-ER 1455

QY 1486 NCNYSDDDAYATAILNPSLSLAVAPDGTIYIADLNIRIRAYSKNPVLAENFOYEA 1545

Db 1456 GDCDFEAHVLATSAKENTIAALAVTPDSHVHIADQANYRIRSVMSSIPASPSREVEY 1515

QY 1546 SPGEQELVFNADGIIHQYTVSLVTGEYLYNFTY---STDNDVTELIDNNGSLKIRDS 1602

Db 1516 APDMQEIYIFNRFCQHVSTRNLTGETTYVFTYVNTSNGKLSVTDAAGKNKVLRLDYT 1575

QY 1603 GMPHLLMPDQIITLTVGTNGGLKVVSTONLEGLMTYDGTGLTATKSDGTGWTTFYD 1662

Db 1576 SQVNSIENTGKQCRRLRTRMKMLHELSTPDNYNVTYEHGPTGLLRTKLDSTGRSVYN 1635

QY 1663 YDHGRLTNVTRPTGVVTSLSHREME-KSITIDI-ENSNRD-----DDVTITNLSSEAS 1715

Db 1636 YDEGRITSAVTPGRIVELSFDLSVKGAVQVKSNAQKEMSLLOIQTATVVRNGAESR 1695

QY 1716 YTVVQDQVRNSYQLCNNGTLRVMYANGMGTISFHSHPVLAGTITPTIGRCNISLPM- 1772

Db 1696 TTVDMD-----GTTTTPWGHNLQMEVAYTYTTLAEQSPLGE---SYVPAKO 1741
QY 1773 -----NGLNSIEWR--LRKE-----QIKG---KVTIFGRKLRVHGRNLLSIDYDRI 1814
Db 1742 RTEIAGDLANREWFYFVRQOPLQAGKQSGKPPRPVTEVGRKLRVNGDNVLTLEYDRET 1801
QY 1815 RTEKIYDHRKRTLRIYDQGRPELWLPSSG-LAAVNVSYFFNGRLAGLQRGAMSERTD 1873
Db 1802 QSVWVMDKQELLNVYDRTSRPISFRPQSGDYADVLDLEYDRFGLSVKMGVLQEAYS 1861
QY 1874 IDKGRIVSRMADGKVNYSYLDK--SNVLLLSQORQ--YIFEYDSSDRLLAVTMPVA 1929
Db 1862 FDRNGLRNLKIDGDSWYAFKMGFSLPLKVTTPRRSDYLLQYDDAGALQSLTPRGH 1921
QY 1930 RHSMSTHTSIGIRNTYNPPESNASVIFDYSDGRLKTSFLTGROVYKYKGLSKLSE 1989
Db 1922 IHAFLQSLGFEKFKQYQYSPINRHFPEILYNDEGQILAKIHPHQSKVAFVHDAGRLT 1981
QY 1990 IYDSTAVTFGYDETGVLLKMYNLQSGFSCITRYR-KIGPLVDKQIYRFSEGMVNARF 2048
Db 1982 ILAGLSHTYVODTTSLVKSVEVEQPGPELRREFKYHAGILKDEKLRFCGKNSLASARY 2041
QY 2049 DTYHDNSFRIASIRPVISETLPVDLYRYDELSGKVEHFGFYIYDI----NQIIT 2104
Db 2042 KYAY-DGNARLSGIEAIDDKELPTTRYKYSONLGOE-----VVQDKITRNAFNRT 2093
QY 2105 AVMTLSKHF-----DTHGRKEVQEMPRSLMYNWTVOYDSMGRVIRKELKLGVPANT 2158
Db 2094 VTQDSKQFFAIVDYDQHGKRVKSLMKNYKIDVFRLELDYDLNRKLSKQTTFGSTAFD 2153
QY 2159 KYTYDYGQGLQSVAVNDPRTPWYSYDLNGN-LHLLNPGNSVRLMPLRYDLRDRITRLG 2217
Db 2154 KINYADG---HVVEVLGNTNKKVLFDENGNTGVVDQGEKEN---LGYDIGDRVIVKG 2206
QY 2218 DVQY-KIDDDGYLCQSGSDIFEYNSKGLLTRYANKA--SGWSVOYRYDGVGRASYKTNL 2274
Db 2207 DVEFNNDYDGFVVRKGEQRYNNRQGLIHSEFERFQSW---YYDDRSRLVAWHDNK 2263
QY 2275 GHHLQYFSDLNPRTHIVYHNHNSSEITSLYLDQGLHFAMESSGEEYVASDNTGTP 2334
Db 2264 GNTQYYVANPRTPHLVTHVHPFKISRTMKLFYDDROMLIALEHED-QRYYYVTTDQNGSP 2322
QY 2335 LAVFINGLMIKQLQYATGEIYYDSNDFQWVIFGHGLYDPLKLVHFTORDYDLVLAG 2394
Db 2323 LAFFDQNGSIVKMERKTPFGRIKIDTKPEFFPIDFHGGLIDPHTKLVYTEQORYDPHVG 2382
QY 2395 RWTSPDYTMKNVNGKEPA-PENLYMFK 2420
Db 2383 QWMTF---LWETLATEMSHTPDVFYIR 2406

RESULT 5
Ti6743
hypothetical protein R13F6.4 - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 17-Mar-2000
C:Accession: Ti6743
R:Miller, N.
submitted to the EMBL Data Library, April 1994
A:Description: The sequence of *C. elegans* cosmid R13F6.
A:Reference number: 218570
A:Accession: Ti6743
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2531 <MIL>
A:Cross-references: EMBL:U00046; NID:q470358; PID:q470361; PIDN:AAC47044.1; GSPDB:GN0002
A:Experimental source: strain Bristol N2; clone R13F6
C:Genetics:
A:Gene: CESP:R13F6.4
A:Map position: 3
A:Introns: 52/2; 113/1; 170/2; 594/1; 1009/1; 1181/1; 1547/3; 1672/3; 1890/1; 2278/1; 24
C:Superfamily: unassigned EGF-related proteins; EGF homology

Query Match 15.1%; Score 2208.5; DB 2; Length 2531;
Best Local Similarity 25.5%; Pred. No. 8.3e-109;
Matches 694; Conservative 441; Mismatches 1008; Indels 581; Gaps 99;
QY 322 TSSGS-TPLFSSSP---GYPLTSGT-----VYTPPRLPLRNTFSRK 360
Db 22 TSGGAPNPTYSDASTLLKYPLAAGTNQNRQRQVGTMMNGDPVAGGPMAL-----SKK 75
QY 361 AFKLKPKSKYCS-WKCAALSAATAAALLA---ILLAYFIVPSLKNSSIDSGEA-----410
Db 76 KKKFDDSDTCSRWPSSKNWILLAAALLVAFICILLFRAPNYVYTPAPSSDSSAAA 135
QY 411 -----EVRRVTQEVPPGVFWRSDIHISQPFKLFNLSLKGDLFGVY 453
Db 136 AASRYQDLGLRALPALISLGERVDVEFFPKSMATYELTVKPSRIFPNATVSGALVLL 195
QY 454 IRRGLPPSHAQYDFM-----ERL-DGKEKWSVVE-----SPRRRSIOTLVON 495
Db 196 MSAGVHPSLSLHDALFPPIRADIRDSKSPTHIVEEFGSRRSRSLGASSSRHRNIELSPR 255
QY 496 EAVFVQYLDVGLHFLAFYNDGDKEMVSENTVVD-----SVQ 533
Db 256 SATFEQFVLEGRHYLTFINERSRVEPISFVAEELQRPPTPTSSSGTGAKEHPLASVL 315
QY 534 DCPNRCHNGGECVSGVCHCFGLGADCAKAAACPVLCGSGQYSGKGTQCQYSGWKGAEC 593
Db 316 VCESNCRQGEVCHGKCHCAPGFTGTCDEAVCPVCSGNGVFGGICVCKSGFKGECE 375
QY 594 VPMNCIDPSCGGHSGC-IDGNCVCSAGYKGHBCEVDCDLPDTCSSHGVGVNCECLSPG 652
Db 376 MRHNCEVADCNRGRCDDTGRRCNPGWTGACELRPHASCHDRGVGVNGTCYCMDG 435
QY 653 WGLNCE-----L 660
Db 436 WGNDCSVFADAIHVHPQAPPRRGQEPTESSKTRKAQVKPTPTSEKKESRELOKPII 495
QY 661 ARVQCPDQ---CSHGTYLPDTGLCSDPNMGPDCSVGCVSDCGTHGVCI-GGACRC 715
Db 496 ATVQVPTESSHPCSAAHQLIDD--ICQCESGWDVDCSQAC--QC-VNGDCLDDGSCQC 550
QY 716 EBGWTAACDQVCHPRCIEHGTG-KDGKCECBEGWHECHTIGRTAGTETDGCPLCN 774
Db 551 WKGWRGSCNCTDKKCAIGCEDRGKASDGSCKCSGNGENCAI-----DGCNOC 601
QY 775 GNGRCTLGONS--WOCVQCTWGRGPCNVAMETSCADNKNDEGDLVDCDLPDCLQAS 832
Db 602 GKGECDMBRRSSENSCRQAGSTGYDCSVSVEMHCDGDLNDSGLIDCDDECCSSSC 661
QY 833 QNSLLCRSRDPLDIIQOQGTDPAPV--KSFYDRIKLLAGKDSHTIIPGENPNSLSVL 890
Db 662 SSESVCSTAASPIELMR----MPPIFNANFAQRVGLIMEKSVQSYTDSQFENENLISV 717
QY 891 IRGOVV-----TTDGTPLGVNVSVFKYKYYTTRDGTGTDLLANGA 935
Db 718 INGRVWGGSGPTGSDDLSTYSNKSSTVPLGVVRVSDAAHPLYGLTTRDGYFDLVNGAR 777
QY 936 SLTLHERAPFMSQERTVWLPWNSFYAMDTLVYMKTEENSIP-----SCDLSGFVR 985
Db 778 SVTLQFLRTQSVKSKSVSPROIHDDIVLYRQSGSPPAISWAPARAKCSPLRRI 837
QY 986 PDPIIISPLSTF-----FSAAPQNPVPEVQVLHEELPGSNVKNLYLSRRTAGYKSL 1041
Db 838 PDVVLISNMQVTSOGIETDETSDSRIVVDSRISFESLPQGTDLVRLVYDSARSAPST 897
QY 1042 LKITWTQSTVPLNLRVHLMVAVEGHLFKQSFQASPNLASTFIWDKTDAYGORVYGLSDA 1101
Db 898 MLIGLYDRVDKELRKVINIRIAGRFRDVLAPRTNITYFAWDMKNAYROSESGLPV 957
QY 1102 VVSVGEFYEYC--PSLLIWEKRTALQGFELDPNLSLGGWSLDKHHILNVKSLHKGTC 1159
Db 958 TVRVGYEQCDRTSERVWQTRRSOMMGATARKMIGTMTWTLDIHHLDIVNNVEMNGG 1017

A:Accession: S14015
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-243,245-369,'V',371-679,'R',681-1676,'I',1678-2139,'TR',2142,2144-2201 <SI
A:Cross-references: EMBL:X56160; NID:g37226; PIDN:CAA39628.1; PID:g37227
R:Nies, D.E.; Hemesath, T.J.; Kim, J.H.; Gulcher, J.R.; Stefansson, K.
J. Biol. Chem. 266, 2818-2823, 1991
A:Title: The complete cDNA sequence of human hexabrachion (tenascin). A multidomain prot
A:Reference number: S16166; MUID:91131572
A:Accession: S16166
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-538,'R',540-1065,'H',1067-1599,'LWHPRASN',1609-2054,'LH',2055-2201 <NIE>
A:Cross-references: EMBL:X56161; NID:g184483; PIDN:AAA8083.1; PID:g184484
R:Glumoff, V.; Savontaus, M.; Vehanen, J.; Vuorio, E.
Biochim. Biophys. Acta 1219, 613-622, 1994
A:Title: Analysis of aggrecan and tenascin gene expression in mouse skeletal tissues by
A:Reference number: S50206; MUID:95035091
A:Accession: S50208
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 46-125 <GLU>
A:Cross-references: EMBL:X80280
Note: the nucleotide sequence was submitted to the EMBL Data Library, July 1994
C:Genetics:
A:Gene: GDB:HXB
A:Cross-references: GDB:120073; OMIM:187380
A:Map position: 9q33-9q33
C:Superfamily: tenascin; EGF homology; fibrinogen beta/gamma homology; fibronectin type
C:Keywords: alternative splicing; extracellular matrix
F:408-434/Domain: EGF homology <EGF>
F:622-703/Domain: fibronectin type III repeat homology <FN3A>
F:711-794/Domain: fibronectin type III repeat homology <FN3B>
F:892-976/Domain: fibronectin type III repeat homology <FN3C>
F:1437-1519/Domain: fibronectin type III repeat homology <FN3D>
F:1619-1701/Domain: fibronectin type III repeat homology <FN3E>
F:1709-1790/Domain: fibronectin type III repeat homology <FN3F>
F:1798-1878/Domain: fibronectin type III repeat homology <FN3G>
F:1886-1966/Domain: fibronectin type III repeat homology <FN3H>
F:1981-2189/Domain: fibrinogen beta/gamma homology <FBG>

Query Match 4.7%; Score 680; DB 2; Length 2201;
Best Local Similarity 20.3%; Pred. No. 1.9e-27;
Matches 454; Conservative 263; Mismatches 789; Indels 726; Gaps 108;
QY 535 CPNCRHNGECYSGVCHCPFGADCAKACPVLCG-GNGQYKSGTCQCYSGWKGACD 593
DB 221 CPSCDNDQKCVNGVCICFEGYAGADCSREICVPVCSEBHGTCDGLCVCHDGFAGDDCN 280
594 VPMNQCIDPSCGGHSCIDGNCVCSAGYKGEHCEVEVDCLDPTCCSHGVCVNGECILCSPGW 653
DB 281 KPL--CLN-NCYNRGRGVNECVCEGTEGTEGDCSELICPN-DCFDRGRGRCINGTCYCEEGF 336
QY 654 GGLNCELARVQPCDSCGHGTYLPDTGLCLSCDPNMGPDSCVEVCSVDCGTHGVCIGGAC 713
DB 337 TGEDC--GRKTPCHACTQGRC--BEGQCVCEGAGFLDCSEKRPADCHNRGRGVDCGR 392
QY 714 RCEEGWTAACDQRVCHPRCIEHGTCCKGKCEKRGWNGEHCITGR-----OTAGTETD 768
DB 393 ECDGFTGADGCELCPCNGCSHGRCVNGCQVCDGEGYTGEDCSQLRCPNDCHSRGRCV 452
QY 769 -----CPDLCNGNRCRTLQNSQWQCQCTGWRGPGCNVAMETSCADNK 811
DB 453 KCVCEQGFQGYCDMSNCPNDCHQHGRCVNG-----MCVCDGTYGTEDCR---DRQCPRDC 505
QY 812 DNEG---DGLVDCLD-----PDC-----CL--QSAQNSLL---CRGSRPL 845
DB 506 SNRGLCVQGVQVCEGFTGPDCAELSCPNCHGQGRVNGCQVCEHGFMGDKCEQRCP 565
QY 846 DTIQGQ--TDPAV--KSFYDRIKLAKGD-STHIIIPENPNSSLVSLIRQVVTIDGT 901
DB 566 DCHGQGRVDCGQCICHEGF-----TGLDCQGHSCPSD---CNLQGVCSRCICNEG- 614

QY 902 PLVGVNVSEVFKYKYGTYTITRODGTDFDLIANGASLTHFERAPFMSQRTVWLPW-NSF 960
DB 615 -YSGEDCSVSPPK-----DLVV-----TEVT-----EETVNLAWDNEM 647
QY 961 YAMDTLVKMTENSIPSCDLSGFRVDPDPIIISPLSTFFSAAPQNPYIPETQVHLHEETE 1020
DB 648 RYTELVVYVTPH-----EGGLEMQF-RVPGD-----QTSTIIQIELE 683
QY 1021 LPSGNVKRLRYLSRTAGYKSLKITMTQSTVPLNLRVHLMVAVEGHLFQKSFQASPNLA 1080
DB 684 -PCVEYFIR-----VFAILENKASIPVSARVATYLPAGECLKFKSKETSVEYE 731
QY 1081 STFIWDKTD-----AYQRYVYGLSDAVVGVFEYE 1110
DB 732 ---NDPLDIAFETWEIIFRNMKDEGEITKSLRRPETSRYQT--GLAP-----GQEYE 780
QY 1111 -----TCPS-----LILWEKRTALLOQFELDPNSL 1135
DB 781 ISLHIVKNNTRPGKLRVTTTRLDAPSQIEVKDVTDTTALITWFKPLAEIDGIEL----- 835
QY 1136 GWSLDKUHILNVKSG--ILHKGKTGENQFLTQOPALITSIMGN--GRRRSISCPSCNGL 1190
DB 836 -----TYGKIDVPGRITIDLTEDENQY-----SIGNALPDTYEYVSLISRRGD 879
QY 1191 AEGNKLLAPVALAVGIDGSLYVGFNYIRRIFFSRNVTISILELRNKEFKHNNPAHKYYL 1250
DB 880 MSSNP--AKETFTTGLDAP-----RNLRRVSGTQNSITL-----EWRNGKAAIDSYRI 925
QY 1251 AVDPVSGSLVSDTNSRRRIYRVKLSGTGDLAGNEVAVAGTGEOLCPDFDEARGDGGKAI 1310
DB 926 KYAPISG-----GDHAEVDVVKSQQA----- 946
QY 1311 DATLMSPRGIADVKNGLMVFYDATMIRKVDQNGIITSLGSDNDLTAVRPLSCDSSMDVAQ 1370
DB 947 -FTKTLTJGL--RPGTEVGIGVSAV-KEDKESNPATINAATELDPDKDLQVSETAETS- 1000
QY 1371 VLEHPTDLA-----VN-----PMDNSLYVLENNVILRTITENHQSIVIAGR 1411
DB 1001 LTLMLKTKPLAKEDRYRLNYSLPTGQWVGVLPRNTTSYVLRG---LEPGQEYNVLLTAEK 1057
QY 1412 PMHCQVPGIDYSLSLKLAHSALESASAIASHTGVL-----YITDE-KKN 1458
DB 1058 GRHKSPARVKASTQQA--PELENLTVEVGDGLRLNWTAAQAYEHFIQVQANKVE 1115
QY 1459 RLRQVTNGEICLLAGASDCDCKNVNCNCYSGDDAYATDALNPSPLAVAPD-GTTY 1517
DB 1116 AARNLTVPGS--LRAVDIPGLKAATPYTVSIXGVIOGYRT-PVLSAEASTGETPNLGEV 1172
QY 1518 IADLGNIRIRAVSKNKPVLNAFNQYEAASPGQEELY---VFNADGIHQVTVSLVTGEYLY 1574
DB 1173 VAEVG-----WDALKLNTWTAPEGAYEYFFIOVEAD-----TVEAAQ 1209
QY 1575 NFETY-----STDNDVTELDNNGNSLK-IRRDSGMPRILLMPDNOILTLTVGTNGGLKV 1628
DB 1210 NLTVPGGLRSLDPLGLKAATHYITIRGVTDFTSTPLSV-----EVLFEVPMGNLTV 1264
QY 1629 ---VSTQNLELGLMTYDNGNTGLLATKSDETGWTTFYDHDREGLTINVTRP----- 1675
DB 1265 TEVSDALRLNWTTPDGTVDQFTIQVQEA-----DQVEEAHLNLTVPGLSRSMELPGL 1316
QY 1676 ---TGVVTSLHREMEKSTIDENSRDDDDVTIT-----NLS 1711
DB 1317 RAGTPTVTLHGEVR-----GHSTRPLAVEVWTEDLPLQGLDLAVSEVGMWGLRLNMTA 1369
QY 1712 VEASVT--VVQDVNSYOLCNN---GTLRVYANGM-----ISFSEPHV 1753
DB 1370 ADNAHEFVQVQEVKVEAAQNLTLPGLSLRAVDIPGLEAATPYRVSYIGVTRGTRTPVL 1429
QY 1754 LAGTIT---PTIGRCNIS-----LPMENGLSIEWRL--- 1782
DB 1430 SAEASTAKEPEIGNLVSDITPESFNLSMWTADGIFETFTIIEIDSNRLLETVEYNISCA 1489

Qy 1783 -RKEQIKG-----VTIFGRKLRVHGRNL-----LSIDYDRN----- 1813
Db 1490 ERTAHISGLPPSTDFIVYLSGLAPSIRKTKISATATTEALPLENLNLS-DINPYGFTVS 1548
Qy 1814 -IRTEKIYDD-----HKKFTLRIIYDOVGRPFLWLPSSGLAAV 1850
Db 1549 WMASENAFDSFLVTVDVSGKLLDPQEPFTLSGTORKLELRGLITGIG-----YEV 1597
Qy 1851 NVSYFFNGRLAGLQGMASERTDDKQGRIVSRMFADGKVMYS-----YLDKSMVLLQOS 1906
Db 1598 MVSFGTGOHQTKPLRAEIVTAEPEVDNLLVSDATPDGFRLSWTADGEVDNFVLKTRDT 1657
Qy 1907 QRQVIFYEYSSDRLLAVTMPVSARHSMSTHTSIGYIRININPPESNASVIFDYSDDGRIL 1966
Db 1658 KKQ-----SEPLEITLLAPERTDLGLREATETEYELYGISKGRRSQTVS-----AI 1705
Qy 1967 KTSPLGTGRQVFFYKGLSKLSEIVDYSTAVTFGYDE-----TTGVLKMWNL 2013
Db 1706 ATTAMGSPKEVIF-----SDITENSATVSWRAPTAQVESFRITYVPIITGGTPSMVT 1757
Qy 2014 QSGGFSCTIRYKRKIGPLVDKQI-----YRFSE-----EGMVNARFDYTYH 2053
Db 1758 D--GTKQTQLRVKLIPGEVLYLSIAKGFEESEPVSGSFTTALDGPGLVTANI----- 1810
Qy 2054 DNSPRIASIKPVISETPLPVDLYRYDEISGKVEHFGKF---GVIIYDINQIITPAVMTLS 2110
Db 1811 TDSEALARMQPAIA---TVDSYVISVTGEKVPEITRTVSGNTVEYALTDLEPATEVTL- 1865
Qy 2111 KHPDTHGRIK-----EVQYEMFRSLMYMTVOYDSMGRVIKREL 2149
Db 1866 RIFAEPQKSSSTTAKFTTDLDSPLDTATEVQSE---TALLTRP----- 1909
Qy 2150 KLGPYANTTKYTYDYG--DGOLQSVAVMNDRTWRYSYDLNGLNHLNPGNSVRLMPLRYD 2208
Db 1910 ---PRASVTGVLVYESVDGTVKEVIGPDTYSVLADLSPSTHY-----TAKIQALNGP 1961
Qy 2209 LRDRI-----TRLGDVQYKIDDDGYLRCORGSDIFEYNSKGLLTRAYN----- 2250
Db 1962 LRSMNIQTIFFTIG--LLYPPFKDCSQAMLNGD----TTSGLTYTLNGLKAQALEVFCDM 2016
Qy 2251 --KASGNSVOVRYDGVGRR-----ASVKNLGHHLQYFYSDLNPNRTHIVYHNSSEI 2302
Db 2017 TSDCGGMIWFLRRKN--GRENFYQNNKAYAGGDRREFFWJGLDNLNKRITTAQGOY----- 2070
Qy 2303 TSLYDYLQGHFLFAMESSEYEEYVASDNTGTPPLAVFSINGLMIR--QLQYATYAGEIYYDSN 2361
2071 -ELRVDLRDH-----GETAFAYDK-----FSVGDAKTRYKLKVEGYSGTAGDS- 2113
Qy 2362 PDFQWVGFHGG 2373
Db 2114 -----MAYHING 2119

RESULT 9
JQ1322
tenascin precursor - mouse
N:Alternate names: contactin; hexabrachion
C:Species: Mus musculus (house mouse)
C:Date: 10-Sep-1999 #sequence.revision 10-Sep-1999 #text_change 16-Jun-2000
C:Accession: JQ1322; A37936; B37936; S14571; S50209
R:Saga, Y.; Tsukamoto, T.; Jing, N.; Kusakabe, M.; Sakakura, T.
Gene 104, 177-185, 1991
A:Title: Murine tenascin: cDNA cloning, structure and temporal expression of isoforms.
A:Reference number: JQ1322; MUID:92009211
A:Accession: JQ1322
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-2019 <SAG>
A:Cross-references: GB:D90343; NID:g220609; PIDN:BAA14355.1; PID:g220610
A:Experimental source: cell line 2H6GR
A:Note: the authors translated the codon ATG for residue 60 as Trp
R:Waller, A.; Beck, S.; Ekblom, P.
J. Cell Biol. 112, 355-362, 1991

A:Title: Amino acid sequence of mouse tenascin and differential expression of two ten
A:Reference number: A37936; MUID:91107734
A:Accession: A37936
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-201,'E',203-317,'S',319-620,622-1010,'N',1012-1018,'S',1020-1024,'H',10
A:Cross-references: GB:X56304
A:Accession: B37936
A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-201,'E',203-317,'S',319-620,622-1010,'N',1012-1018,'S',1020-1024,'H',10
A:Cross-references: GB:X56304
R:Waller, A.; Beck, S.; Ekblom, P.
submitted to the EMBL Data Library, August 1990
A:Description: Amino acid sequence of mouse tenascin and differential expression of tw
A:Reference number: S14571
A:Accession: S14571
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-201,'E',203-317,'S',319-620,622-1010,'N',1012-1018,'S',1307-2019
A:Cross-references: EMBL:X56304; NID:g54768; PIDN:CAA39751.1; PID:g54769
R:Glumoff, V.; Savontaus, M.; Vehanen, J.; Vuorio, E.
Biochim. Biophys. Acta 1219, 613-622, 1994
A:Title: Analysis of aggrecan and tenascin gene expression in mouse skeletal tissues
A:Reference number: S50206; MUID:95035091
A:Accession: S50209
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: mRNA
A:Residues: 46-146 <GLU>
A:Cross-references: EMBL:X80281
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, July 1994
C:Superfamily: tenascin; EGF homology; fibrinogen beta/gamma homology; fibronectin ty
C:Keywords: alternative splicing; calcium binding; cell adhesion; duplication; extrac
F:1-22/Domain: signal sequence #status predicted <SIG>
F:23-2019/Product: tenascin, long splice form #status predicted <MAT>
F:23-1071.1527-2019/Product: tenascin, short splice form #status predicted <MAT2>
F:408-434/Domain: EGF homology <EGF>
F:622-703/Domain: fibronectin type III repeat homology <FN3A>
F:711-793/Domain: fibronectin type III repeat homology <FN3B>
F:802-884/Domain: fibronectin type III repeat homology <FN3C>
F:892-976/Domain: fibronectin type III repeat homology <FN3D>
F:984-1064/Domain: fibronectin type III repeat homology <FN3E>
F:1073-1155/Domain: fibronectin type III repeat homology <FN3F>
F:1346-1428/Domain: fibronectin type III repeat homology <FN3G>
F:1437-1519/Domain: fibronectin type III repeat homology <FN3H>
F:1527-1608/Domain: fibronectin type III repeat homology <FN3I>
F:1616-1696/Domain: fibronectin type III repeat homology <FN3J>
F:1704-1784/Domain: fibronectin type III repeat homology <FN3K>
F:1799-2007/Domain: fibrinogen beta/gamma homology <F8G>
F:38,166,184,327,788,1018,1079,1093,1119,1184,1210,1275,1301,1354,1364,1394,1627,1878

Query Match 4.7% Score 679.5; DB 1; Length 2019;
Best Local Similarity 20.3%; Pred. No. 1.7e-27;
Matches 421; Conservative 251; Mismatches 686; Indels 715; Gaps 98;
Qy 535 CPRNCHNGECVSGVCHCFPGFLGADCAKACPVLCISNGQYSGKTCQCYSGWKGAEC-D 593
Db 346 CPNDGCGQCEEGCQCVNEGFACDCSEKRCPADCHRRGRLNGQCEDDGFTGADGD 405
Qy 594 VPMNQCIDPSCGGHSGCIDGNCVCSAGYKGBEHCBEVDCLDPTCSSHGVCVNGECLCSPGW 653
Db 406 L---QCPN-GCSGHGRVNGQCVDEGYTGDBDCSQRRCPN-DCHNRGLICVQKGCICEQGF 460
Qy 654 GGLNCELARVQCPDQCSGHGTYLPDTGLCSDPNMWDPCSVESVDCGTHGVICGAC 713
Db 461 KGFDG---SEMSCPNDCHQHGRV---NGMCICDDDTGDECDRDRRCPRDCSQRGRVGDQC 516
Qy 714 RCEEGWTGAACDQVRVCHPRCTEHCCTCKDGKCECEGNGEHCITGRQTAGTGTGCGPDLG 773
Db 517 ICEDGFPGDCAELSCPSDCHGRCVNGQCICHEGFTGDKCEQR-----CFSDC 567
Qy 774 NNGRCTLGQNSWOCVQOTGWRGPGCNVAMETSCADNKDNEGDGLVDCLDPCCLQSACQ 833

Db 568 HGQGRCEG---QCICHGFTGLDGC---QRSCPNDCSNOGO-----CVSGRC- 609
QY 834 NSLLCRGRDPLDIIOGGOTDMPAVKSFYDRIKLAGKDSHIIPGENPNSLSVLIRG 893
Db 610 -----ICNEGVT----- 616
QY 894 QVVTDTGLVGNVSVFYKPKYGYTITRODGTFDLIANGASLTLHFRAPFMSQERTV 953
Db 617 -----GIDCSEVSPK-----DLIV---TEVT-----EEIV 639
QY 954 WLPW-NSFVMTDLVNMKTENS-----IPSCDLSGFRPDPPIIISPLSTFFSAARG 1004
Db 640 NLAWDNEMRVETYLIMYTHADGLEMQFRVPGDQSTTIR-----ELEPG 685
QY 1005 ONPIPETOVLEHIEILPGSNVRLVSSRTAGYKSLKITWTQSVPLNLRVHLMAV 1064
Db 686 VEYFIRVFAILENKRSP-----VSARVATY-----LPAP 715
QY 1065 EGHLFQKSFQASPLASTFIWKTDAYQORVYGLSDAVSVGFYETPCPSLILWE----- 1119
Db 716 EGLKPKSIRKETSVEVE---WDPLD-----IAFE-----TWEIFERN 748
1120 -----KRTALLOQFELDPNSLGGHSLDKHHIL--NVKSGILHKGT 1157
Db 749 MNKEDEGITKSLRRPETSYROTGLAPGOEYIS-----LHTVKNTRGPGKLVY 799
QY 1158 G-----ENQFLTQQAITSIMNGRRRSISCPSCNGLAEGNKKLAPALAVGI-- 1206
Db 800 TTRLDAPSHIEVKDVTDTTALTWF-----KPLAE-----IDSIELSGIKD 841
QY 1207 -DGLSYGDFNYIRIRFPGRNVTISLE---LRNKEFKHNNPAHKYLL-ADVPVSGSLY 1260
Db 842 VPGDRTTIDLTHEDNOYSIGNLRPDPEYEVSLISRRVDMASNAKETFITGLDAPNLR 901
QY 1261 VSDTNS-----RRV---SLSGTKDLAGNSEVAVAGTGEQCLPDEARCGDG 1306
Db 902 VSQDONSITLERNVKNADISYRIKVIAPISG---GDHAEIDVPKSOQA----- 946
QY 1307 GRAIDATLMSPRGIAVDKNGLMVFVDATMIRKVDONGIISTLLGSNDLPAVRPLS-CDSS 1365
Db 947 -----TTKTTLGL---RGTGYGICVSAY-KGDRESDPATTNAATEIDAPKDLRSETT 997
QY 1366 MDVAQVRLEWPTDLAVNPMDSLYLVLENNVILRITENHQVSIAGRMHCQV----- 1417
Db 998 QD---SLTFFWTPLA--KFD-----RYRLNY--SLPTGOSMEVQLPKDATSHV 1039
QY 1418 ----PGIDYLSKLAIHSALESASAIASHGTGLYITETDEKKINLRQVTTNGEICLL 1472
Db 1040 LTDLEPGQEYT-----VLLIAEKGRHKSPARVKASTAEVPSL 1077
1473 AGAASDCCKNDVNCYSGDDAYATDAI-----LNSPSSLAVAPDGTIYIA 1519
Db 1078 ENLTVTEAGWDGLRLNWTADDLAYEFVLOVEANNVETAHNETVPGNLRRAADIPGLKVA 1137
QY 1520 DLGNIRIRAVSK--NKPVLNAFNQVEAASPGQEQLYVFNADGHQVTVSLVTEYLYN-- 1575
Db 1138 TSYRVSIIYGARYRTPVLSA-ETSGTTPNLGEVTVAEV-GWDALTLNMTAPEGAYKNF 1195
QY 1576 FIYSTDNVTELDIDNNGNSLKRROSSGMPRHLLMPDNOIITLVCTNGGLKVVSTONLE 1635
Db 1196 FIOVLEADTITQVQN-----LTV--PGGLRSVDLPGLK 1226
QY 1636 LGLMTVDGNTGLLATKSDGTGWTTFYDDHEGRLTNVTRPTGVVTSLHREMEKSIITDIE 1695
Db 1227 AA-----TRYI-----TLRGVTDQFGTA-----PLSVEV- 1251
QY 1696 NSNRDDVTVIINLSVSREASYT-----VVQDVQRNSVOLCNRN---GTL 1735
Db 1252 ----LTEDLPOLGGLSVTEVSWDGLTLNMTTDDLAYKHFFVQVOEANNVAAQNLTVPVGL 1308
QY 1736 RVNYANGMG-----ISPHSEPHVLAGTIT---PTIGRCNLSLPMENGLNSIEW 1780

Db 1309 RAVDIPGLKADTPYRVSYIYQGYRTPMLSTDVSTAREPEIGNLNVSDVTPKSFN-LSW 1367
QY 1781 RLURKEQIKGVITFRGKLRVHGRNLLSIDYDRNIRTEKIYDDHRKFTLRIIYDQVGRPEL 1840
Db 1368 ----TATDGIIFDMFTIEI-----IDSNRLLOTAENINISGAERTAH- 1404
QY 1841 WLPSSGL-AAVNVSYPFNGLRGLAQRGAMSERTDIDKQRI---VSRMFADCKVMSYS- 1894
Db 1405 ----SGLPPSTDFIVYLSGIAPSIKTIISTATTAEPEVDNLLVSDATPDGFRLSWTA 1460
QY 1895 ---YLDKSMVLLQSORQYIFEDSSDRLLAVTMPVSARHSMSTHTSICGIYRINIYPPES 1951
Db 1461 DEGIFDSFVIRIDRTKKQ-----SEPOBISLPSPERTDITGLREATEYEIYGISRG 1514
QY 1952 NASVIFYDSDGRILKTSFLGTGROVEYKYGKLSKLSSEIYVDYSTAVTFG----- 2000
Db 1515 RR-----SOPVSATATTAMGSPKEIMF-----SDITENNATVSWRAPTAQVESFRI 1560
QY 2001 --YDETTGVLKVMYNSQSGFSCTIRYRKIGPLVDKQIYRES-----EE----- 2041
Db 1561 TVVPMTGGAPSMVTVD--CTDTETRLVKLTGVE---YRVSVIAMKGFESDPVSGTLIT 1615
QY 2042 -----GMYNARFDYTHDHSFRIASIKPVISETPLPVDLY-----RYDEISGKVEHF 2088
Db 1616 ALDGPSEGLLIANI-----TDSEALAMWQPAIA---TVDSYVISYTGERVPEVTRTVSG- 1665
QY 2089 GRFGVIYDINQIITFAVMTLSKHFDTHGRK-----EVOYEMF 2127
Db 1666 ---NIVEYELHLEPATEYELS-IFAEKGOQKSSITATKFTTDLDSPREFTATEVQSE-- 1719
QY 2128 RSLMYMTVOYDSMGKRVIKRELKLPYANTKYTYDYDG-DGOLQSVAVNDRPTWRYSYD 2186
Db 1720 TALLTWRP-----PRASVTGYLLVYESVDGTGVEIVGPDPTTSYSLAD 1762
QY 2187 LAGNLHLLNPGNSVRLMPLRYDLRORI---TRLDGVQYKIDDDGYLCORG-----S 2234
Db 1763 LSPSTHY-----SARIQALSGSLRSKLIQTIIFTIG-LLYPPFRD---CSQAMLNGDTS 1813
QY 2235 DIFEVNSKGLLTRY-----NKASGMSVOYRYDCVGR--ASAKYTLNGLHLOQYF 2281
Db 1814 GLYTIYINGDKTOALEVYCDMTSDGGWIVFLRRKN-GREDFYRNWKAAGAAGDRREF 1872
QY 2282 YSDLINPTIRITHVYNNHNSSEITSLYDLOGLHFAMESSGEYEVASDNTGTPLAVFSIN 2341
Db 1873 WLGLONLSKITAQQOV-----ELRVDLQDH-----GESAYAVYDR-----FSVG 1911
QY 2342 GLMIK-OLQYATAYGEIYYDSNPDFQWVIGFHGS 2373
Db 1912 DAKSRYTKLAKVEGYSGTAGDS-----MNYHNG 1937
RESULT 10
S19694
tenascin precursor - plg
N:Alternate names: contactin; hexabrachion
C:Species: Sus scrofa domestica (domestic pig)
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: S19694
R:Nishi, T.; Weinstein, J.; Gillespie, W.M.; Paulson, J.C.
Eur. J. Biochem. 202, 643-648, 1991
A:Title: Complete primary structure of porcine tenascin. Detection of tenascin transcr
A:Reference number: S19694; MUID:92104189
A:Accession: S19694
A:Molecule type: mRNA
A:Residues: 1-1746 <NIS>
A:Cross-references: EMBL:X61599; NID:g2124; PIDN:CAA43796.1; PID:g2125
C:Superfamily: tenascin; EGF homology; fibrinogen beta/gamma homology; fibronectin ty
C:Keywords: alternative splicing; calcium binding; cell adhesion; duplication; extrac
F:1-22/Domain: signal sequence #status predicted <SIG>
F:23-1746/Product: tenascin #status predicted <NAT>
F:346-372/Domain: EGF homology <EGF>
F:372-403/Domain: EGF homology <EGF2>
F:622-703/Domain: fibronectin type III repeat homology <FN3A>

535	CPNRCHNGCEVSGVCHCFPGFGLGADCAKAACPVLCGSGNQYYSKGTQCQCYSGWKGAECVD	594
315	CPKDCDFRGRCIINCTYCDGFEDEGDCRLACPHGCRGRGRCBEGQCVCEGDEGAGDCS-	373
595	PMNOCIDPS--CGHGSGCIDNGCVCSAGYKGEHCHEEVDCLDPTCSSHGVCNNGECLSPGM	653
374	-ERRC--PSCDHNRCRLDGRCECDGFEDEGDCGEURC--PGGCSGHRGVCNQCVCDBER	429
654	GGINCELARVQCDQCSGHHGTYPDPTGLCSDPNMNGPDCSEVCSVDCTGHGVCIGGAC	713
430	TGEDC--SOLRCPNDCHGRGCV--QGRCEHGFQGYDCSEMSCPHDCHQHGRCVNGMC	485
714	RCEEGWTGAACDORVCHPRCIENGHTCKDGKCECRBGWNGEHCTIGRQTAGTETDGGPDLC	773
486	VCDDGYTGEDRELRCPGDCSQBRGCVDRGCVCEHGFAGPDCA-----DLACPSC	536
774	NGNRCTLQGNQWCQVCOTGWRGPGCNVAMETSCADNKONEGDLVDCLDPOC-----	826
537	HGRGRVNG---QCVCHEGFTGKDCG---QRRCPG--DCHGOG--RCVDGQCVCHGEFT	585
827	---CLQSACONSLLCRGSRDPLDIIOGGTDPWPAVXFYDRIKLAGKOSTHIIPGENPF	883
586	GLDCGQRSCFN-----DCSNMGQ-----	603
884	NSSLVSLIRGOVVTDTGLVGVNVSFVKYPKYGTITRQDGTFDLIANGGASLTUHFER	943
604	-----CVSGRCINEG--YSGEDCSQVSPK-----DLIV--TEXT-----	635
944	APFWSQERTVLPW-NSFYAMDTLVNMKTENS-----IPCDLSGFRVDPDPIIISPP	994
636	-----EETVNLAWDNEMRVTELYIVVPTPHEDGLEMQFRVPGDQTSTIR-----	680
995	LSTFFSAAGONPIVPTOVHLHHEELPGSNVLYRVLSSRTAGYKSLKAITWTQSTVPLN	1051
681	-----ELEPGVEFIRVFAILENKKSI-----VSARVAY-----	711
1055	LIRVHLMAVVEGHLFOKSFQASPNLASTFTWKTDAYGQRYGLSDAVVSVGFYEYTCPS	1111
712	-----LPTPEGLKFKSIKETSVEE---WDPLD-----IAFE-----	740
1115	LILWE-----KRTALOGFELDSPNLGGWSLOKHHLNWK	1141
741	--TWEIFIRNNHKEDEGEITKSLRRPETYRTQGLAPGQYEYS-----LHI--VK	787
1150	SGILHKTGTENOFITQPAITISMGNGRRSSICPS--CNCGLAENKLLA--PVALA	1201
788	NNT--RCPGLKRVTTTR-----LDAPSQTEAKDVTDTTALITWFKPLAEI	830
1204	VGDGSLYVGDFNYIRRIFFPSNRVTSILELRNKEFKHSNNPAHYLAVDPVSGSLYSD	1261
831	DGIELTYGKD-----VFGDRTT--IDLTHEENQYS-----IGNL-KPD	866
1264	TNSRRIRVKSLSGTKDLACNSEVWAGTGBQCLPFDEARCGCGGKAIDATLMSPRGIADV	1321
867	TE-----YEVSLISRRADMSN-----PAKE-----TFTTGLDAPRN-----	898

C>Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jan-2000
C:Accession: T09070
R:Rowen, L.; Mahairas, G.; Qin, S.; Ahearn, M.E.; Dankers, C.; Lasky, S.; Loretz, C.; S. submitted to the EMBL Data Library, October 1997
A:Description: Sequence of the mouse major histocompatibility locus class III region.
A:Reference number: 216543
A:Accession: T09070
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-4006 <ROW>
A:Cross-references: EMBL:AF030001; NID:g2564945; PID:g2564958
C:Genetics: TNX
A:Gene: TNX
A:Map position: 17
A:Introns: 124/1; 735/1; 773/3; 826/1; 914/1; 1037/1; 1135/1; 1329/1; 1440/1; 1501/1; 3115/1; 3208/1; 3302/1; 3405/1; 3517/1; 3558/1; 3606/1; 3694/1; 3737/3; 3819/1; 3915/1; 4006/1
C:Superfamily: tenascin-X; EGF homology; fibrinogen beta/gamma homology; fibronectin type I; fibronectin type II; fibronectin type III
C:Keywords: extracellular matrix
F:422-448/Domain: EGF homology <EGF>
F:826-906/Domain: fibronectin type III repeat homology <3FR>
F:3789-3997/Domain: fibrinogen beta/gamma homology <FBG>

Query Match 4.3%; Score 632.5; DB 2; Length 4006;
Best Local Similarity 24.0%; Pred. No. 1.7e-24;
Matches 253; Conservative 108; Mismatches 358; Indels 337; Gaps 49;

QY 531 SYVDCPRNCHGNGCVSGVCHYCPGFLGADCAKACPVLCGNGQVSKGTCOCYSGMKA 590
DB 418 STRACPRDCRGRCRGEDGVCVCHAGYSGEDCGVRCPGDCRGNCESGRVCWFCYGR 477
QY 591 ECDVPMNQIDPSCGGHSCIDGNCVCSAGYKGEHEEYDCLDPTCSSHGVNCGELCS 650
DB 478 DCG---TRACPGDCRGRCVDCVCPNPGFTGDCGSRRC-PGDCRGHGHGCVCA 533
QY 651 PGWGLNCLARVQCPDQCSHGTYLPDTGLCSDBNMGWPCSVESVDCGTHGVCIG 710
DB 534 VGYSGDCSTR--SCPSDCRGRCQCL--NLGCECEGYSGEDCGIRRCPRDCSQHGVCO 589
QY 711 GACRCEEGWTGAACQQRVCHPRCIEHGTCKGCKCEGNGEHCITIGRTAGTDTGCP 770
DB 590 GLCMCHAGYAGEDCSRTCPADCRRCRGCEGRCVCPNGYTGACA-----TRTCP 640
QY 771 DLGNGNGRCTLGQ-----NSQVCVQTCQWRGPG 798
DB 641 ADCRGRCVQVCMCYGVYSGEDCGQEPASACPGGCPRELCRAGQCVCEGFRGP 700
QY 799 CNVAMETSCADKNEGDLVCLDPCCLQACQNSLLCR-----GSRDP----- 844
DB 701 C--AIQT--CPG-----DCRSRGECIQRC-----VCQEGYAGDDCGEIPAIQNM 743
QY 845 LDIIQOG--QTDW-----PAYKSFYDRIKLLAGKD- 872
DB 744 MHLLETTVTEWTRAPGPDVAYEIOFIPMTSGVSPPTARVPSSASAYDQRLAPGQDY 803
QY 873 --STHIIGEN--PFNSLSVLIRG---QVVTDTGT-----PLGVNVSFVKTPK 915
DB 804 QVTVALRGTSPPASKTITTMIDGQDLRVAVVPTTLDLSWLRPOAEVDRFVVSYS 863
QY 916 VGYTTR-----QDGT--FDLIANGASLTJLHFERAPMSOERTVWLPWNSFYAMD 967
DB 864 AGNQRVLEVPPEADQTLTDMPGVEYVVTVAERGHVASYPASI----- 909
QY 968 MKTEENSIPCDLSGFVRPDPPIISPLSTFFSAAPGNPIVPETQVLHEETELPGSNVK 1027
DB 910 -RANTGSLPSGLLEATDEPPP--SGPSTQGAQAP---ILLLEHPLGE----- 952
QY 1028 LRYLSRRTAGYKSLKIMTQSTVPLNLRVHLMVA-----VEGHLFOKSOQASPN 1078
DB 953 LKVLGRDKAG---RLSVAMTAQPSFAHQLRMQVAEGPWAHEELLPGDVQQAALVPPPPP 1009
QY 1079 LASTFIWDKTDAGORVYGLS--DAVSVGVFEYETC-----PSLILWEKRTALLQG 1127

DB 1010 GA-----PYKULFLHGITPGGKISVPIVYOGIMDRAQEQGPKPSV---QPRLG----- 1053
QY 1128 FELDPNSLGGSLDKHHLNPKSGILHGTGENQ---FLTQO-----PAITTSINGNRR 1179
DB 1034 -ELTVTGLTSDSL-----LLHWTVPGEFDSFLIQYKDKGQPAI-SVEGPQRS 1100
QY 1180 RSISCPSCNGLAEGNK-----LLAPVALAVGID-----GSLYVGD 1214
DB 1101 TPIS-----GLEPGRKYFILYGLIGKKRHGPLMAEAKILIQSDPDGSPRRLGELWTD 1155
QY 1215 -----FNVIIRIFPSPNNVTSILELNKEKFKHNNNAHKYYLAVDPVSG---SLYSDTNS 1266
DB 1156 PTPHSLHLSWTVLGQDFSVVQYRDKE-----GQP-----RVVPVEGDRKSWISPLDP 1205
QY 1267 RRIYRVKLSGTLKDLAGNSEVAVGTGEQCLPDEARCGDGGKAIDATLMSPRGIADVKN 1326
DB 1206 NRKYRF-----TLF---GIANKKRY 1222
QY 1327 LMYFVDATMIRKVDQNGIISTLLGSNDLTAVRPLSCDSMDVAQVRLDPTDLAVNPM 1386
DB 1223 GPLTADGTTAAETKESESSEPPRLGELTVGTVP---DS-----URLSW--TVARGPFD 1271
QY 1387 SLVVLNNVILLRITENHQVSI-IAGRPMHCQVPGID 1421
DB 1272 FV-----ILYKDAQGQPSQVPIEGDENEVTVPGLE 1301

RESULT 12
T12457
hypotheical protein DKFp56400423.1 - human (fragment)
C:Species: Homo sapiens (man)
C>Date: 23-Jul-1999 #sequence_revision 23-Jul-1999 #text_change 23-Jul-1999
C:Accession: T12457
R:Wambutt, R.; Heubner, D.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S. submitted to the Protein Sequence Database, June 1999
A:Reference number: 217524
A:Accession: T12457
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-184 <WAM>
A:Cross-references: EMBL:AL080120
A:Experimental source: fetal brain; clone DKFp56400423
C:Genetics:
A:Note: DKFp56400423.1

Query Match 4.3%; Score 627.5; DB 2; Length 184;
Best Local Similarity 62.4%; Pred. No. 1.8e-26;
Matches 116; Conservative 39; Mismatches 28; Indels 3; Gaps 1;

QY 2548 ASEDSRKVASVLNNAYLDKMHYSIEGKDTYFVKIGSADGLVTLGTTIGRKVLES 2607
DB 2 ANEDGRRVAAILDHAYLENLHFTIDGVDTHYFVKPGPSEGDLATLGLSGRRRTLENG 61
QY 2608 VTVSPTLLVNGTRFRFTNIEFOYSTLLSIRYGLTPTDLEEKARVLDQARALGTAW 2667
DB 62 VTVSINTVLNGRTRYTDIQLQYGCALCNTRYG---TTLDEEKARVLELQARVROAW 118
QY 2668 AREQOKARDREGSLWTEGEGKQQLLSTGRVQGYGVYVLPVEQYPELADSSNIFQLRQ 2727
DB 119 AREQORLEGEGLRAWTEGEGKQQLSTGRVQGYGVYVLPVEQYPELADSSNIFQLRQ 178
QY 2728 NEMGKR 2733
DB 179 SEMGRR 184

RESULT 13
A40701
tenascin-X precursor - human
C:Species: Homo sapiens (man)
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Dec-1999
C:Accession: A40701; A33725; C42175

Db 1411 APEPHGELTVEATGHTLH-LSMVTEGEEDSFEIQYTRDQLOMVRIGGRNDITLS 1469
QY 1489 CYSGDDAYATDALNSPSSILAVAPDCTIYIADLGNIRIRAVSKN-----KPVLNAP 1539
Db 1470 GLESDHRYLV-----FLYGFSDG-----KHVGPVHVHVALTEPPATPEPIKPRUGEL 1517
QY 1540 NOYEAAAPGEQEELYVFNADGIIHQYTVSLVTGEYLYNFTYSTONDVPELIDNNGSLKIR- 1598
Db 1518 TVTID-ATPDSLSL-----SWTVP--EGGF-----DHFLVQYRNGDGQPKAVRG 1557
QY 1599 -RDSSOMPRLHMPDNQIITLTVGTNGG-----LKVSTQNLE----- 1635
Db 1558 PGHEGVTISGLEPDHKYKMNLYGFHGGQRMGPVSVVGTAAEETPSPTEPSMEAPEA 1617
QY 1636 ---LGLMYDGNLGLATKSDGTGT-----TFYDYDHEGRLTNVTRPTG----- 1677
Db 1618 EEPGLGELVTGS-----SPDSLSLSTVPGQRFSDSTVQYKORDGR-PQVVRVGGPEEES 1672
QY 1678 -----VVTSLHREMEKSIITDIENSRRDDVTITNLSSVEASYTVVQDVQRNSYQLCNN 1732
Db 1673 PDAPLAKRLGOMTVRDITSD-----SLSLSWTVPPEGOFDH----- 1708
QY 1733 GTLRVNYANGMISFSEPHVL-----AGTITPTIGRCNISLPMENGLNSIEHRLK 1784
Db 1709 ---FLVQFKNGDG-----QPKALRVPGHEDGABEETPS-----PTEPSMEAPE--PPE 1751
QY 1785 EQIKGKVTIFG 1795
Db 1752 EPLGLGELVTG 1762

RESULT 14

A43902
tenascin - eastern newt (fragment)
C:Species: Notophthalmus viridescens, Triturus viridescens (eastern newt)
C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 12-Feb-1999
C:Accession: A43902
R:Onda, H.; Poulin, M.L.; Tassava, R.A.; Chiu, I.M.
Dev. Biol. 148, 219-232, 1991
A:Title: Characterization of a new tenascin cDNA and localization of tenascin mRNA during gastrulation in the embryo of the eastern newt.
A:Reference number: A43902; MUID:92038434
A:Accession: A43902
A:Molecule type: mRNA
A:Residues: 1-647 <OND>
A:Cross-references: GB:M76615
A:Note: sequence extracted from NCBI backbone (NCBI:64543, NCBI:64547)
C:Superfamily: tenascin; EGF homology; fibronogen beta/gamma homology; fibronectin type I
C:Keywords: extracellular matrix; glycoprotein; tandem repeat
F:287-313/Domain: EGF homology <EGF>
F:316-427/Domain: fibronectin type III repeat homology <FN1>
F:435-517/Domain: fibronectin type III repeat homology <FN2>

Query Match 4.1%; Score 604; DB 2; Length 647;
Best Local Similarity 28.2%; Pred. No. 2.7e-24;
Matches 153; Conservative 75; Mismatches 207; Indels 108; Gaps 23;

QY 535 CPNCHNGECVSGVCHCPFGFLGADCAKAAACPVLCSSGNGQYSGKTCQCYSGWKGAECV 594
Db 70 CPNDCFDRCRNGVCFCDGEGTGEDCGELTCPNCCNRRGRCVNGLCVCDGDFGQDDCS- 128
QY 595 PMNQCTDPSCGGHSGICDNGCVCSAGYKGEHCEEVDCLDPTCSSHGVCVNGECLSPGNG 654
Db 129 -ELRCPN-DCNDRGRGVNGKVCKEGFMGEDCADLRCPN-DCNNRGRGVNGQVCVDEGFM 185
QY 655 GUNCELARVQCDCSGHCTYLPDTGLCSLCPNPMGPDSCSVEVCSVDCGTHGVCITGGACR 714
Db 186 GEDC--SDLRCPGDCNNRGRGV--NGQVCVDEGFMGEDCGELRCPPDCNNRGRGVNGQCI 241
QY 715 CEEGWTGAACDORVCHPRICIEHGTCKDGKCEGREGNWEHCTIGRTAGTETDGGCPDLCN 774
Db 242 CDEGFMGNGCELRCPCNDCKNRGRCVNGQICDDGFKGEDCSLR-----CPDDCN 292

QY 775 GNGRCTLGQNSQCYCQTGWRGPGCNVAMETSCADNKNEDGLVDCLDPDCCL--QSAC 832
Db 293 DRGRCLNG---QCVAEGFTGEND---SLACLN-----CNDRGCLVNGQCVC 335
QY 833 QNSLL---CRGRDPLDIIIOQGTOWMPAVKSPYDRIKLL-----AKDSTHIIP 878
Db 336 EGFGLGEDCSEVSPKDLTVDTVTQSVNLEWANEKVTLEYLITYPTSPGGLDLDFRVP 395
QY 879 GENPNSLSVLIRQOVVTTDGTPLGVNVSVFKYKPYCTITRODGTDLTANGASL- 937
Db 396 GDOT-TATIQLEPG-----VEYFVRVFAIRNQRSIPVSARVATHLP 437
QY 938 ---TLHFERAPMSOERTVWLPMN---SFYAMDITLVMT-BENSTPSCDLSGFVRPDI 989
Db 438 TTDDLRFKSV---KETSVEVEMDPLDISFDTWLIIINTKEENGEISLSQ--RP--- 487
QY 990 IISPLSTFFSAAPQNPVLPETQVLHBEIPLPGSNVKLYLSSRTAGYKSLKLTMTQS 1049
Db 488 -VTSYVQT--GLAPGET-----YNFSIHVVKNSTRG-PGLAKVTTRL 526
QY 1050 TVP 1052
Db 527 DAP 529

RESULT 15

T42629
tenascin-X - bovine
N:Alternate names: flexilin
C:Species: Bos primigenius taurus (cattle)
C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 21-Jul-2000
C:Accession: T42629
R:Elefteriou, F.; Exposito, J.Y.; Carrone, R.; Lethias, C.
J. Biol. Chem. 272, 22866-22874, 1997
A:Title: Characterization of the bovine tenascin-X.
A:Reference number: Z22180; MUID:97426436
A:Accession: T42629
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-4135 <ELE>
A:Cross-references: EMBL:Y11915; NID:g2462978; PIDN:CAA72671.1; PID:g2462979
C:Gene: TN-X
C:Superfamily: tenascin-X; EGF homology; fibronogen beta/gamma homology; fibronectin
C:Keywords: extracellular matrix; glycoprotein; heptad repeat

Query Match 4.0%; Score 590; DB 2; Length 4135;
Best Local Similarity 26.4%; Pred. No. 3.4e-22;
Matches 172; Conservative 63; Mismatches 216; Indels 200; Gaps 27;

QY 532 VODCPNCHNGECVSGVCHCPFGFLGADCAKAAACPVLCSSGNGQYSGKTCQCYSGWKGA 591
Db 400 VRSCPSDCNQRCEGRCVCHPFGYSGPDCGARACPRDCRGRGRCNGVCHVCHAGYSGED 459
QY 592 CDV-----PMNQCI-----DPSCGGHSGICDNGCVCSAGYK 623
Db 460 CGVRSCPGDCRRRRGRGRCGRCVCPGTYGTRACPDGDCRGRGRCVDCRCVCPNPFAG 519
QY 624 EHCEEVDCLDPTCSSHGVCVNGECLSPGNGGLNCELARVQCDCSGHGTYLPDTGLCS 683
Db 520 EDCGSRRC-PGDCRGRGRCGDCVCDVGYEGEDC--GKRSCPRCQGRGQCL--EGRCV 574
QY 684 CDPNMGPDSCSVEVCSVDCGTHGVCIGGACRCEEGWTGAACDORVCHPRICIEHGTCKDGK 743
Db 575 CDDGYEGEDCGVRRCPRCNQRGVCDGVCTCWEFGAGEDGLRVCPNCHRRGRGRCNGR 634
QY 744 CECREGNWEHCTIGRTAGTETDGCPLDCLNGNGRCTLGQNSWQCVQCTGWRGPCN--- 800
Db 635 CVCDSGYTGPSCA-----TRTCPADCRGRGRCVQG---VCVCHVGYSGEDCGQEE 681
QY 801 ---VAMETSCADNKNEDGLVDCLD----PDCCLQSAQNSLLCRG---SRDPLDIQQ 850

```
Db 682 PPASACPGCGP-RELCSAGCVCVEGFRGPDCAIQT-CFGD--CRGRGECREGSCVQD 737
Qy 851 G-----QTDW-----855
Db 738 GYAGEDCGEVPALIEGHRMHLLETTVTRAPGNVDAYEIQFIPTTEGASPPFTARV 797
Qy 856 PAVKSFYDRIKLAGKD---STH1PGEN---PNSLSVSLIRG-----QVTTDGT---- 901
Db 798 PSSASAYDQRLAPGQEVQVTVRALRGTNMGPPASKTITTMIDGPDQLRVAVVPTTLEL 857
Qy 902 -----PLGVNVVSFYKYPKYGTITR-----QDGTG--DLIANGCASLTLHFERAPENSQ 949
Db 858 NWLRPQAEVDREVVYSVSAGNQVRLEVPSADGTLTGLMPGVEYVYVTVTAERGRAVSY 917
Qy 950 ERTVWLPWNFSFYAMDTLVMKTEENSIPSCDLSGFVRPDP11ISSPLSTFFSAAP--GONP 1007
918 PASIRA--NTGSSLSGLLGATDE-----PPP---SGPSTTQGAQAPVLOQRP 959
1008 -----IVPETQVVLHEEIE LPG 1023
Db 960 QELAE LRVLGKDKTGR LRVAVTAQPDFTTHFQLRLRVPEGPGAHEEL-LPG 1009
```

Search completed: September 18, 2002, 11:09:55
Job time: 265 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 18, 2002, 11:08:25 ; Search time 23.68 Seconds
(without alignments)
4468.772 Million cell updates/sec

Title: US-09-800-198-8

Perfect score: 14581

Sequence: 1 MDVKDRHRSLTRGCRGEC.....ELADSSNQFLRQNMGRK 2733

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

ched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	732.5	5.0	1808	1 TENA_CHICK	P10039 gallus gall
2	680	4.7	2201	1 TENA_HUMAN	P24821 homo sapien
3	656	4.5	1746	1 TENA_PIG	Q29116 sus scrofa
4	616	4.2	4289	1 TENX_HUMAN	P22105 homo sapien
5	463.5	3.2	2531	1 NTC1_MOUSE	Q01705 mus musculus
6	462.5	3.2	2531	1 NTC1_RAT	Q07008 rattus norv
7	450.5	3.1	2318	1 NTC3_MOUSE	Q61982 mus musculus
8	448.5	3.1	2444	1 NTC1_HUMAN	P46531 homo sapien
9	437.5	3.0	1064	1 FBPI_STRPU	P10079 strongyloe
10	428.5	2.9	2524	1 NOTC_XENLA	P21783 xenopus lae
11	417	2.9	2139	1 CRB_DROME	P10040 drosophila
12	417	2.9	2437	1 NOTC_BRARE	P46530 brachydanio
13	415.5	2.8	830	1 SREC_HUMAN	Q14162 homo sapien
14	413	2.8	2703	1 NOTC_DROME	P07207 drosophila
15	399	2.7	1295	1 GLP1_CABEL	P13508 caenorhabdi
16	396.5	2.7	1376	1 CRBH_HUMAN	P82279 homo sapien
17	395.5	2.7	1408	1 SERR_DROME	P18168 drosophila
18	395	2.7	833	1 DL_DROME	P10041 drosophila
19	391	2.7	1964	1 NTC4_MOUSE	P31695 mus musculus
20	368.5	2.5	473	1 FP2_MYTGA	Q25464 mytilus gal
21	363	2.5	686	1 DLL4_MOUSE	Q31711 mus musculus
22	360.5	2.5	1429	1 LI12_CABEL	P14585 caenorhabdi
23	350	2.4	570	1 FBP3_STRPU	P49013 strongyloe
24	346.5	2.4	714	1 L1L1_RAT	P97677 rattus norv
25	345	2.4	379	1 WIF1_MOUSE	Q89411 mus musculus
26	344	2.4	379	1 WIF1_HUMAN	Q39545 homo sapien
27	343	2.4	723	1 DLL1_HUMAN	Q00548 homo sapien
28	338.5	2.3	685	1 DLL4_HUMAN	Q9nr61 homo sapien
29	335.5	2.3	3075	1 LMA1_HUMAN	P25391 homo sapien
30	332	2.3	722	1 DLL1_MOUSE	Q61483 mus musculus
31	328.5	2.3	4660	1 LRP2_RAT	P98158 rattus norv
32	327	2.2	3672	1 LML2_CABEL	Q21313 caenorhabdi
33	322.5	2.2	2334	1 WAPA_BACSU	Q07833 bacillus su

RESULT 1

ID	TENA_CHICK	STANDARD:	PRT:	1808 AA.
AC	P10039: P13132: 073584; 073585;			
DT	01-MAR-1989 (Rel. 10, Created)			
DT	01-MAR-1992 (Rel. 21, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Tenascin precursor (TN) (Hexabrachion) (Cytotactin) (Neuronectin) (GWMEM) (JI) (Mitotendinous antigen) (Glioma-associated-extracellular matrix antigen) (GP 150-225).			
DE	Gallus gallus (Chicken).			
OS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.			
OC	NCBI_TaxID=9031;			
OX	[1]			
RN	SEQUENCE FROM N.A.			
RP	TISSUE=Embryo;			
RC	MEDLINE=90030407; PubMed=2478295;			
RX	Spring J., Beck K., Chiquet-Ehrismann R.;			
RA	"Two contrary functions of tenascin: dissection of the active sites by recombinant tenascin fragments.";			
RT	Cell 59:325-334(1989).			
RL	[2]			
RN	SEQUENCE OF 27-722 FROM N.A., AND SEQUENCE OF 79-96.			
RP	TISSUE=Fibroblast;			
RC	MEDLINE=89030589; PubMed=2460335;			
RX	Pearson C.A., Pearson D., Shibahara S., Hofsteenge J., Chiquet-Ehrismann R.;			
RA	"Tenascin: cDNA cloning and induction by TGF-beta.";			
RT	EMBO J. 7:2977-2982(1988).			
RL	[3]			
RN	SEQUENCE OF 464-1018 AND 1412-1661 FROM N.A., AND SEQUENCE OF 852-868.			
RP	TISSUE=Embryo;			
RC	MEDLINE=88176910; PubMed=2451243;			
RX	Jones F.S., Burgoon M.P., Hoffman S., Crossin K.L., Cunningham B.A., Edelman G.M.;			
RA	"A cDNA clone for cytotactin contains sequences similar to epidermal growth factor-like repeats and segments of fibronectin and fibronogen.";			
RT	Proc. Natl. Acad. Sci. U.S.A. 85:2186-2190(1988).			
RL	[4]			
CC	- FUNCTION: SAM (SUBSTRATE-ADHESION MOLECULE) THAT APPEARS TO INHIBIT CELL MIGRATION. MAY PLAY A ROLE IN SUPPORTING THE GROWTH OF EPITHELIAL TUMORS. IS A LIGAND FOR INTEGRINS ALPHA-8/BETA-1, ALPHA-9/BETA-1, ALPHA-V/BETA-3 AND ALPHA-V/BETA-6.			
CC	- SUBUNIT: HEXAMERIC. A HOMOTRIMER MAY BE FORMED IN THE TRIPLE COILED-COIL REGION AND MAY BE STABILIZED BY DISULFIDE RINGS AT BOTH ENDS. TWO OF SUCH HALF-HEXABRACHIONS MAY BE DISULFIDE LINKED WITHIN THE CENTRAL GLOBULE.			
CC	- SUBCELLULAR LOCATION: Extracellular matrix.			
CC	- ALTERNATIVE PRODUCTS: 3 ISOFORMS; 230 KDA (SHOWN HERE), 200 KDA AND 190 KDA; ARE PRODUCED BY ALTERNATIVE SPLICING. THEY ARE PRODUCED IN A TISSUE-AND TIME-SPECIFIC MANNER DURING DEVELOPMENT.			
CC	- INDUCTION: BY TGF-BETA.			
CC	- SIMILARITY: CONTAINS 13.5 EGF-LIKE DOMAINS.			
CC	- SIMILARITY: CONTAINS 11 FIBRONECTIN TYPE III-LIKE DOMAINS.			

088516 mus musculus
P98164 homo sapien
O88671 rattus norv
Q9ny77 homo sapien
Q07954 homo sapien
P24043 homo sapien
Q00918 rattus norv
Q00174 drosophila
P19137 mus musculus
Q61554 mus musculus
Q61001 mus musculus
P35555 homo sapien

ALIGNMENTS

34 312 2.1 592 1 DLL3_MOUSE
35 312 2.1 4655 1 LRP2_HUMAN
36 307.5 2.1 589 1 DLL3_RAT
37 302.5 2.1 618 1 DLL3_HUMAN
38 297 2.0 4544 1 LRP1_HUMAN
39 296 2.0 3110 1 LMA2_HUMAN
40 295.5 2.0 1712 1 TGFB_RAT
41 294.5 2.0 3712 1 LMA_DROME
42 291 2.0 3084 1 LMA1_MOUSE
43 288.5 2.0 2871 1 FBNI_MOUSE
44 287.5 2.0 3718 1 LMA5_MOUSE
45 284.5 2.0 2871 1 FBNI_HUMAN

RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-95155442; PubMed-7531707;
 RA Gherzi R., Carnemolla B., Siri A., Ponassi M., Balza E., Zardi L.;
 RT "Human tenascin gene. Structure of the 5'-region, identification, and
 RT characterization of the transcription regulatory sequences.";
 RL J. Biol. Chem. 270:3429-3434(1995).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-91131572; PubMed-1704365;
 RA Nies D.E., Hemesath T.J., Kim J.H., Gulcher J.R., Stefansson K.;
 RT "The complete cDNA sequence of human hexabrachion (tenascin). A
 RT multidomain protein containing unique epidermal growth factor
 RT repeats.";
 RL J. Biol. Chem. 266:2818-2823(1991).
 RN [4]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-92052108; PubMed-1719530;
 RA Gulcher J.R., Nies D.E., Alexakos M.J., Ravikant N.A., Sturgill M.E.,
 RA Marton L.S., Stefansson K.;
 RT "Structure of the human hexabrachion (tenascin) gene.";
 RT Proc. Natl. Acad. Sci. U.S.A. 88:9438-9442(1991).
 RN [5]
 RP SEQUENCE OF 431-2055 FROM N.A. (ISOFORM P31).
 RC TISSUE-Glioblastoma;
 RX MEDLINE-89160821; PubMed-2466295;
 RA Gulcher J.R., Nies D.E., Marton L.S., Stefansson K.;
 RT "An alternatively spliced region of the human hexabrachion contains a
 RT repeat of potential N-glycosylation sites.";
 RL Proc. Natl. Acad. Sci. U.S.A. 86:1588-1592(1989).
 RN [6]
 RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF FIBRONECTIN TYPE-III 3.
 RX MEDLINE-93068293; PubMed-1279805;
 RA Leahy D.J., Hendrickson W.A., Aukhil I., Erickson H.P.;
 RT "Structure of a fibronectin type III domain from tenascin phased by
 RT MAD analysis of the selenomethionyl protein.";
 RL Science 258:987-991(1992).
 CC -|- FUNCTION: SAM (SUBSTRATE-ADHESION MOLECULE) THAT APPEARS TO
 CC INHIBIT CELL MIGRATION. MAY PLAY A ROLE IN SUPPORTING THE GROWTH
 CC OF EPITHELIAL TUMORS. IS A LIGAND FOR INTEGRINS ALPHA-8/BETA-1,
 CC ALPHA-9/BETA-1, ALPHA-V/BETA-3 AND ALPHA-V/BETA-6.
 CC -|- SUBUNIT: HEXAMERIC. A HOMOTRIMER MAY BE FORMED IN THE TRIPLE
 CC COILED-COIL REGION AND MAY BE STABILIZED BY DISULFIDE RINGS AT
 CC BOTH ENDS. TWO OF SUCH HALF-HEXABRACHIONS MAY BE DISULFIDE LINKED
 CC WITHIN THE CENTRAL GLOBULE.
 CC -|- SUBCELLULAR LOCATION: Extracellular matrix.
 CC -|- ALTERNATIVE PRODUCTS: FOUR VARIANTS ARE PRODUCED FROM A SINGLE
 CC GENE IN A TISSUE- AND TIME-SPECIFIC MANNER DURING DEVELOPMENT.
 CC -|- INDUCTION: BY TGF-BETA.
 CC -|- SIMILARITY: CONTAINS 15 EGF-LIKE DOMAINS.
 CC -|- SIMILARITY: CONTAINS 15 FIBRONECTIN TYPE III-LIKE DOMAINS.
 CC -|- SIMILARITY: CONTAINS 1 FIBRINOGEN C-TERMINAL DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X56160; CAA39628.1; -;
 DR EMBL; X78565; CAA55309.1; -;
 DR EMBL; M55618; AAA88083.1; -;
 DR EMBL; M24630; AAA52703.1; -;
 DR PIR; S14015; S14015.
 DR PDB; 1TEN; 31-OCT-93.
 DR MIM; 187380; -;
 DR InterPro; IPR000361; EGF-like.
 DR InterPro; IPR003961; FN.III.
 DR InterPro; IPR002181; Fibrinogen_C.
 DR Pfam; PF00008; EGF; 14.
 DR Pfam; PF00147; fibrinogen_C; 1.

DR PFam; PF00041; fn3; 15.
 DR SMART; SM00181; EGF; 9.
 DR SMART; SM00001; EGF_Like; 1.
 DR SMART; SM00186; FBG; 1.
 DR SMART; SM00060; FN3; 12.
 DR PROSITE; PS00022; EGF_1; 14.
 DR PROSITE; PS01186; EGF_2; 15.
 KW Glycoprotein; Cell adhesion; Repeat; EGF-like domain; Coiled coil;
 KW Extracellular matrix; Signal; Alternative splicing; 3D-structure.
 FT SIGNAL 1 22
 FT CHAIN 23 2201
 FT DOMAIN 118 145
 FT DOMAIN 174 186
 FT DOMAIN 186 217
 FT DOMAIN 217 248
 FT DOMAIN 248 280
 FT DOMAIN 280 311
 FT DOMAIN 311 342
 FT DOMAIN 342 373
 FT DOMAIN 373 404
 FT DOMAIN 404 435
 FT DOMAIN 435 466
 FT DOMAIN 466 497
 FT DOMAIN 497 528
 FT DOMAIN 528 559
 FT DOMAIN 559 590
 FT DOMAIN 590 621
 FT DOMAIN 621 710
 FT DOMAIN 711 801
 FT DOMAIN 802 891
 FT DOMAIN 892 983
 FT DOMAIN 984 1071
 FT DOMAIN 1072 1162
 FT DOMAIN 1163 1253
 FT DOMAIN 1254 1344
 FT DOMAIN 1345 1435
 FT DOMAIN 1436 1526
 FT DOMAIN 1527 1617
 FT DOMAIN 1618 1708
 FT DOMAIN 1709 1797
 FT DOMAIN 1798 1885
 FT DOMAIN 1886 1973
 FT DOMAIN 1982 2201
 FT DISULFID 190 200
 FT DISULFID 194 205
 FT DISULFID 207 216
 FT DISULFID 221 231
 FT DISULFID 225 236
 FT DISULFID 238 247
 FT DISULFID 252 263
 FT DISULFID 256 268
 FT DISULFID 270 279
 FT DISULFID 284 294
 FT DISULFID 288 299
 FT DISULFID 301 310
 FT DISULFID 315 325
 FT DISULFID 319 330
 FT DISULFID 332 341
 FT DISULFID 346 356
 FT DISULFID 350 361
 FT DISULFID 363 372
 FT DISULFID 377 387
 FT DISULFID 381 392
 FT DISULFID 394 403
 FT DISULFID 408 418
 FT DISULFID 412 423
 FT DISULFID 425 434
 FT DISULFID 439 449
 FT DISULFID 443 454
 FT DISULFID 456 465
 FT DISULFID 470 480
 FT DISULFID 474 485
 FT DISULFID 487 496

PFam; PF00041; fn3; 15.
 SMART; SM00181; EGF; 9.
 SMART; SM00001; EGF_Like; 1.
 SMART; SM00186; FBG; 1.
 SMART; SM00060; FN3; 12.
 PROSITE; PS00022; EGF_1; 14.
 PROSITE; PS01186; EGF_2; 15.
 Glycoprotein; Cell adhesion; Repeat; EGF-like domain; Coiled coil;
 Extracellular matrix; Signal; Alternative splicing; 3D-structure.
 SIGNAL 1 22
 CHAIN 23 2201
 DOMAIN 118 145
 DOMAIN 174 186
 DOMAIN 186 217
 DOMAIN 217 248
 DOMAIN 248 280
 DOMAIN 280 311
 DOMAIN 311 342
 DOMAIN 342 373
 DOMAIN 373 404
 DOMAIN 404 435
 DOMAIN 435 466
 DOMAIN 466 497
 DOMAIN 497 528
 DOMAIN 528 559
 DOMAIN 559 590
 DOMAIN 590 621
 DOMAIN 621 710
 DOMAIN 711 801
 DOMAIN 802 891
 DOMAIN 892 983
 DOMAIN 984 1071
 DOMAIN 1072 1162
 DOMAIN 1163 1253
 DOMAIN 1254 1344
 DOMAIN 1345 1435
 DOMAIN 1436 1526
 DOMAIN 1527 1617
 DOMAIN 1618 1708
 DOMAIN 1709 1797
 DOMAIN 1798 1885
 DOMAIN 1886 1973
 DOMAIN 1982 2201
 DISULFID 190 200
 DISULFID 194 205
 DISULFID 207 216
 DISULFID 221 231
 DISULFID 225 236
 DISULFID 238 247
 DISULFID 252 263
 DISULFID 256 268
 DISULFID 270 279
 DISULFID 284 294
 DISULFID 288 299
 DISULFID 301 310
 DISULFID 315 325
 DISULFID 319 330
 DISULFID 332 341
 DISULFID 346 356
 DISULFID 350 361
 DISULFID 363 372
 DISULFID 377 387
 DISULFID 381 392
 DISULFID 394 403
 DISULFID 408 418
 DISULFID 412 423
 DISULFID 425 434
 DISULFID 439 449
 DISULFID 443 454
 DISULFID 456 465
 DISULFID 470 480
 DISULFID 474 485
 DISULFID 487 496

FT DISULFID 501 BY SIMILARITY. 511
FT DISULFID 505 BY SIMILARITY. 516
FT DISULFID 518 BY SIMILARITY. 527
FT DISULFID 532 BY SIMILARITY. 542
FT DISULFID 536 BY SIMILARITY. 547
FT DISULFID 549 BY SIMILARITY. 558
FT DISULFID 563 BY SIMILARITY. 573
FT DISULFID 567 BY SIMILARITY. 578
FT DISULFID 580 BY SIMILARITY. 589
FT DISULFID 594 BY SIMILARITY. 604
FT DISULFID 598 BY SIMILARITY. 609
FT DISULFID 611 BY SIMILARITY. 620
FT CARBOHYD 38 N-LINKED (GLCNAC. .) (POTENTIAL). 38
FT CARBOHYD 166 N-LINKED (GLCNAC. .) (POTENTIAL). 166
FT CARBOHYD 184 N-LINKED (GLCNAC. .) (POTENTIAL). 184
FT CARBOHYD 327 N-LINKED (GLCNAC. .) (POTENTIAL). 327
FT CARBOHYD 788 N-LINKED (GLCNAC. .) (POTENTIAL). 788
FT CARBOHYD 1018 N-LINKED (GLCNAC. .) (POTENTIAL). 1018
FT CARBOHYD 1034 N-LINKED (GLCNAC. .) (POTENTIAL). 1034
FT CARBOHYD 1079 N-LINKED (GLCNAC. .) (POTENTIAL). 1079
FT CARBOHYD 1093 N-LINKED (GLCNAC. .) (POTENTIAL). 1093
FT CARBOHYD 1119 N-LINKED (GLCNAC. .) (POTENTIAL). 1119
FT CARBOHYD 1184 N-LINKED (GLCNAC. .) (POTENTIAL). 1184
FT CARBOHYD 1210 N-LINKED (GLCNAC. .) (POTENTIAL). 1210
FT CARBOHYD 1261 N-LINKED (GLCNAC. .) (POTENTIAL). 1261
FT CARBOHYD 1275 N-LINKED (GLCNAC. .) (POTENTIAL). 1275
FT CARBOHYD 1301 N-LINKED (GLCNAC. .) (POTENTIAL). 1301
FT CARBOHYD 1366 N-LINKED (GLCNAC. .) (POTENTIAL). 1366
FT CARBOHYD 1392 N-LINKED (GLCNAC. .) (POTENTIAL). 1392
FT CARBOHYD 1445 N-LINKED (GLCNAC. .) (POTENTIAL). 1445
FT CARBOHYD 1455 N-LINKED (GLCNAC. .) (POTENTIAL). 1455

Query Match 4.7%; Score 680; DB 1; Length 2201;
Best Local Similarity 20.3%; Pred. No. 6.2e-29;
Matches 454; Conservative 263; Mismatches 789; Indels 726; Gaps 108;

QY 535 CPRNCHNGECVSGVCHFCFPGFLGADCAKAAACPVLCG-NGOYKSGTCQCYSGHWKGAEC 593
DB 221 CPSCDQGGKCVNGVGCFCFEGYAGADCSREICPVPCSEHGTCVDGLCVCHDGFAGDDCN 280

QY 594 VPMNQCIDPSCGGHSGIDGNCVCSAGYKGEHCBEVCDLPDTCSSHGVCVNGECLCSPGW 653
DB 281 KPL--CLN-NCYNRGRVENCVECDGFTGEDCSLLCPN-DCFDRCINGTCYCEBGP 336

QY 654 GGLNCELARVQCPDQCSGHGTYLPDTGLCSDCPNMWPGDCSVEYCVSDCGTHGVCIGAC 713
DB 337 TGEDC--GKPTCPHACTQGRG--EEGQCVCDGFAGLDCSEKRCPADCHNRGRVCVDRG 392

QY 714 RCEEGWTGAACDQVCHPRCIEHGTCKDKGCEGREGNHECHTIGR-----QTAGTETDG 768
DB 393 ECDGFTGADCGELKCPNGCSGHGRVNGQCVCDGTYGDCSOLRCPNDCHSRGRVCG 452

QY 769 -----CPDLCNGRGTCLGONSMQCVCTGWRGPGCNVAMETSCADNK 811
DB 453 KVCQEGFKGYDCSDMPCNDCHQGRVNG-----MCVCDGTYGEDCR---DRQCPDRC 505

QY 812 DNEG--DGLVDCLD---PDC-----CL--QSAQNSLL---CRGSRDL 845
DB 506 SNRGLCVGQCVCDGFTGPDCAELSCPNCHGQGRVNGQCVCHGFGMKDCKEQRCPS 565

QY 846 DIIQOQO-TQMPAV--KSFYDRIKLAGKD-STHIFGENPNSLSVSLINGQVTTDGT 901
DB 566 DCHGQGRVDCQICHEGF-----TGLDGOHQSCPSD---CNNLQCVSGRGCICNEG- 614

QY 902 PLVGVNVFVKYKGYTIHQDGTFDLIANGGASLTLHFERAFPMQERTVWLPW-NSF 960
DB 615 -YSGEDCSEVSPK-----DLVW---TEVT-----EETVNLAWDNEM 647

QY 961 YAMDTLVNKTENSIPSCDLSGFRVDPPIIISPLSTFFSAAPGONPIVPTQVLHHEIE 1020
DB 648 RVTEYLVVYTPTH-----EGGLEMQF-RVPGD-----QTSTIIIELE 683

QY 1021 LPSGNKRLYLSRTAGYKSLKTKITMTQSTVPLNLRVHLMAVEGHFQKSFQASPNLA 1080

DB 684 -PGVEYFIR-----VFALENKSPVSARVATVLPAGEGLKFKSIKETSVEVE 731
QY 1081 STFIWDKTD-----AYQORVYGLSDAVSVGEYE 1110
DB 732 ----WDPDLIAFETWEIIFRNMKKEDEGETKSURRSETSRYOT--GLAP-----GOEYE 780
QY 1111 -----TCPS-----LILWEKRTALQGFELDPNSL 1135
DB 781 ISLHIVKNNTRPGCLKRVTTITRLDAPSQIEVKQDVTDITALITWFKPLAEIDGIEL----- 835
QY 1136 GWSLDKHHILNVKSG--ILHKGTEGNOFLTQOAPITSMGN---GRRRSISPCSNGL 1190
DB 836 -----TYGKIDVPGDRTTIDLTEDENQY-----SIGNLKPDTEVSVISRRGD 879
QY 1191 AEGNKLAPVALAVGIDGSLVVGDFNVRIRFPSSNNVTSILELRNKFKNHNPAAHYL 1250
DB 880 MSSNP--AKETFTGLDAP-----RNLRRVSQTDNSITL-----EWRNGKAAIDSVRI 925
QY 1251 AVDPVSGSLVSDTNSRRIRYRVKSLSGTKDLAGNSEYVAGTGEQCLPFDEARCGDGGKAI 1310
DB 926 KYAPISG-----GDHAEDVDVPSQQA----- 946
QY 1311 DATLMSPRGIADVKNGLMYFVDATMIRKVDONGIISTLLSNDLTAVRPLSCDSSMDVAQ 1370
DB 947 -TTKTTLTGL---RPGTEYGIGVSAY-KEDKESNPATINAATELTDTPKDLQVSETAETS- 1000
QY 1371 VRLEWPTDLA-----VN-----PMDNSLYVLNNVILRTTENHOVSIIAGR 1411
DB 1001 LTLWKTPPLAKFDYRLNYSLPTQWVGVLPRNTTSYVLRG---LEPGQBYNVLTLAEK 1057
QY 1412 PMHQVPCIDYLSKLAHSAESAIAISHTGVL-----YITETDE-KKIN 1458
DB 1058 GRHKSPARVKASPEQA--PELENLTVEVGWDLRLNWTAAOQAYEHFIQVQANKVE 1115
QY 1459 RLROVTTNGEICLLAGASDCDCKDNVNCYSGDDAYATDAIINSPLSLAVAPD-GTIY 1517
DB 1116 AARNLTVPGS--LRADVPLGLKAATPYTSYIGYQGYRT--PVLSEASTGETPNLGEW 1172
QY 1518 IADLGNIRIRAVSNKPNVNAFYEAASPGQEELY---VFENADGIIHOYTYSVLTGVEVLY 1574
DB 1173 VAEVG-----WDALKLNTAPEGAYEYFFIQVEAD-----TVEAAQ 1209

QY 1575 NFTY-----STDNDVTELIDNNGSLK-IRRDSSGMRPHLLMPDNQIITLTVGNGGLKV 1628
DB 1210 NLTVPGGLRSTDLPLGLKAATHYITIRGVTQDFSTPLSV-----EVLTEVPDNGNLTV 1264

QY 1629 --VSTQNLLELCLMTYDGNLTGLLATKSDETGWTTFYDYDHEGRLTNTRP----- 1675
DB 1265 TEVSWDALRLNWTTPDGTQDFTIQVQEA-----DQVEEAHLNTPVGSLSRSMETPGL 1316

QY 1676 ---TGVTYSLHREMEKSTIDIENSRRDDDVIT-----NLSS 1711
DB 1317 RAGTPYTVTLHGEVR-----GHSRPLAVEVVTEDLPQLGLADVSEVGHGDLRLNWT 1369

QY 1712 VEASYT--VVQDQVRNSYQLCNN---GTLRMVYANGM-----ISFSEPHV 1753
DB 1370 ADNAYEHFVIOQVQEVKNVVEAAQNLTLPGSLRAVDIPGLEAATPYRVSYIGYRTPVL 1429

QY 1754 LAGTIT--PTIGRCNIS-----LPMENGLANSIEWRL--- 1782
DB 1430 SAEASTAKEPEIGNLVNSDITPESFNLSMATDGFETFTTEIIDSNRLLLETVEYNISGA 1489

QY 1783 -RKEQIKGK-----VTIFGRKLVRHGRNL-----LSIDYDRN----- 1813
DB 1490 ERTAHISGLPPSTDFIVYLSGLAPSIRTKTISATATTEALPPLLENLTIS-DINYPGFTVS 1548

QY 1814 -IRTEKIYDD-----HRKFTLRIYDQVGRPFLMLPSSGLAAV 1850
DB 1549 WMASENADFSLTVVDVDSKLLDPQEFTLSCGTQRKLERLGLITGIG-----YEV 1597

QY 1851 NVSYFFNRLAGLAGRAMSERTIDKQGRIVSRMFADCKWVSYS-----YLDKSWLLQOS 1906

Db 1598 MVSFGTQGHQTKPLRAEIVTEAPEVDNLLVSDATPDGFLRLSWTADGVEFVFLKIRDT 1657
QY 1907 QRQYIFEYDSSDRLLAVTPSVAHSMSTHTSIGYIRNYNPESNASVIFDYSDDGRIIL 1966
Db 1658 KQO-----SEPLEITLLAPETRLDTGLREATEVEIEIYGLSKGRSQTVS-----AI 1705
QY 1967 KTSFLGTGRQYKYKGLSKLSEIYVDVSTAVTFGYDE-----TTGVLKMWNL 2013
Db 1706 ATTAMGSPKEVIF-----SDITENSATVSWRAPTAQVESFRITVYVITGTPGSMVTV 1757
QY 2014 QSGGSCITRYRKIGPLVDKQI-----YRFE-----EGWVAREDYTH 2053
Db 1758 D--GKTKQTRKLKIPGVEIYLSIIAMKGFEESEPVSGFTTALDGPGLVTANI----- 1810
QY 2054 DNSFRIASIKPVISPTPLVDLYRDEISGKVEHFGKE---GVIVYDINQIITTAVMTLS 2110
Db 1811 TQSEALARKQPAIA---TVDSYVVISYTGKVEPTRTVSGNTVEYALTDLEPATETYL- 1865
QY 2111 KIFDTHGRIK-----EVQYEMFRLMYWMTVQVDSMGRVTKREL 2149
Db 1866 RIFAEBKGPQKSTITAKFTTDLSDPRDLTATEVQSE--TALLTWRP----- 1909
2150 KLGPMANTTKYDYDG--DQOLQSVAVNDPRTWRYSYDLNGNLHLLNPGNSVRLMPLRYD 2208
Db 1910 ---PRASVTGYLLVYESVDGTVKEVIGVDPDTSYSLADUSPTHY-----TAKIQALNGP 1961
QY 2209 LRDR1-----TRLGDVQYKIDDDGYLCQSGDIFEYNSKGLLTRYVN----- 2250
Db 1962 LRSNMQITFTTIG--LLYFPKDCSQAMLNGD---TTSGLYTIYLVGDKAQALEVFCDM 2016
QY 2251 ---KASGWSVQYRVDGVR--ASQYTNLGHILQYPSYDLNPNTRTHVYHNHNSSEI 2302
Db 2017 TSDGGGWIYFLRKRN--GRNFYQNKAYAAAGFGDRFEFLWGLDNLNKITAQGY----- 2070
QY 2303 TSLYVDLQGLHFAMESSGEEYVVASDNTGTGPLAYSINGLMK--OLOVYAYGEIYDGN 2361
Db 2071 -ELRVDLRH-----GETAFAYDK-----FSVGDAKTRYKLKRVGYSGTAGDS- 2113
QY 2362 PDFQMVIGPHGG 2373
Db 2114 -----MAYHNG 2119
RESULT 3
TENA_PIG STANDARD; PRT: 1746 AA.
AC Q29116; P98142;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
Tenaascin precursor (TN) (Hexabrachion) (Cytotactin) (Neuronectin)
(GMBM) (Ji) (Miotendinous antigen) (Glioma-associated-extracellular
matrix antigen) (GP 150-225) (Tenaascin-C) (TN-C) (P230).
GN HXB.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Submaxillary gland;
RX MEDLINE=92104189; PubMed=1722152;
RA Nishi T., Weinstein J., Gillespie W.M., Paulson J.C.;
RT "Complete primary structure of porcine tenascin: detection of
tenascin transcript in adult submaxillary glands.";
RL Eur. J. Biochem. 202:643-648(1991).
RN [2]
RP SEQUENCE OF 813-825; 887-917; 998-1011; 1597-1608 AND 1719-1730.
RC TISSUE-Fetal brain;
RX MEDLINE=98158323; PubMed=9498558;
RA Wakatsuki S., Ho S.H., Arioka M., Yamasaki M., Kitamoto K.;
RT "Isolation and characterization of a 230 kDa protein (p230)
specifically expressed in fetal brains: its involvement in neurite

outgrowth from rat cerebral cortex neurons grown on monolayer of
astrocytes.";
J. Biochem. 122:1146-1152(1997).
CC FUNCTION: SAM (SUBSTRATE-ADHESION MOLECULE) THAT APPEARS TO
INHIBIT CELL MIGRATION. MAY PLAY A ROLE IN SUPPORTING THE GROWTH
OF EPITHELIAL TUMORS. IS A LIGAND FOR INTEGRINS ALPHA-8/BETA-1,
ALPHA-9/BETA-1, ALPHA-V/BETA-3 AND ALPHA-V/BETA-6.
CC FUNCTION: PLAYS A ROLE DURING EARLY BRAIN DEVELOPMENT PARTICULARLY
IN GROWTH CONE GUIDANCE. INVOLVED IN NEURITE OUTGROWTH FROM
CORTICAL NEURONS GROWN ON THE MONOLAYER OF ASTROCYTES.
CC SUBUNIT: HEXAMERIC. A HOMOTRIMER MAY BE FORMED IN THE TRIPLE
COILED-COIL REGION AND MAY BE STABILIZED BY DISULFIDE RINGS AT
BOTH ENDS. TWO OF SUCH HALF-HEXABRACHIONS MAY BE DISULFIDE LINKED
WITHIN THE CENTRAL GLOBULE.
CC SUBCELLULAR LOCATION: Extracellular matrix.
CC ALTERNATIVE PRODUCTS: 3 ISOFORMS: MAJOR, MINOR-1 AND MINOR-2
(SHOWN HERE); ARE PRODUCED BY ALTERNATIVE SPLICING. THEY ARE
PRODUCED IN A TISSUE-AND TIME-SPECIFIC MANNER DURING DEVELOPMENT.
CC TISSUE SPECIFICITY: SUBMAXILLARY GLANDS AND BRAIN.
CC DEVELOPMENTAL STAGE: PREDOMINANTLY EXPRESSED IN THE EMBRYONIC AND
EARLY POSTNATAL STAGES. LITTLE OR NO DETECTION IN ADULT BRAIN.
CC INDUCTION: BY TGF-BETA.
CC SIMILARITY: CONTAINS 15 EGF-LIKE DOMAINS.
CC SIMILARITY: CONTAINS 11 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC SIMILARITY: CONTAINS 1 FIBRONECTIN C-TERMINAL DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
EMBL: X61599; CAA43796.1; -
HSSP: P24821; ITEN.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR003961; FN_III.
DR InterPro: IPR002181; Fibrinogen_C.
DR InterPro: IPR003962; FnIII_repeat.
DR InterPro: IPR002049; Laminin_EGF.
DR Pfam: PF00008; EGF_14.
DR Pfam: PF00147; fibrinogen_C; 1.
DR Pfam: PF00041; fn3; 10.
DR PRINTS: PR00011; EGF_LAMININ.
DR PRINTS: PR00014; FNTYPEIII.
DR SMART: SM00181; EGF; 9.
DR SMART: SM00001; EGF_Like; 2.
DR SMART: SM00186; FBG; 1.
DR SMART: SM00060; FN3; 9.
DR PROSITE: PS00022; EGF_1; 15.
DR PROSITE: PS01186; EGF_2; 14.
KW Glycoprotein; Cell adhesion; Repeat; EGF-like domain; Coiled coil;
KW Extracellular matrix; Alternative splicing; Signal.
FT SIGNAL 1 22
FT CHAIN 23 1746
FT DOMAIN 23 185 INVOLVED IN HEXAMER FORMATION.
FT DOMAIN 118 145 COILED COIL (POTENTIAL).
FT DOMAIN 174 186 EGF-LIKE 1 (INCOMPLETE).
FT DOMAIN 187 217 EGF-LIKE 2.
FT DOMAIN 218 249 EGF-LIKE 3.
FT DOMAIN 250 280 EGF-LIKE 4.
FT DOMAIN 281 311 EGF-LIKE 5.
FT DOMAIN 312 342 EGF-LIKE 6.
FT DOMAIN 343 373 EGF-LIKE 7.
FT DOMAIN 374 404 EGF-LIKE 8.
FT DOMAIN 405 435 EGF-LIKE 9.
FT DOMAIN 436 466 EGF-LIKE 10.
FT DOMAIN 467 497 EGF-LIKE 11.
FT DOMAIN 498 528 EGF-LIKE 12.
FT DOMAIN 529 559 EGF-LIKE 13.
FT DOMAIN 560 589 EGF-LIKE 14.
FT DOMAIN 590 620 EGF-LIKE 15.


```
Db 1066 --RVKASTAGEPE-----IGNLSVSDITPE---SFLSMTATEGAFETFTIEIDSN 1112
QY 1592 GNSLKIRRSSGNPR--HL--LMPDQNIITLVGTNGGL--KVVS-----TONLE 1635
Db 1113 RLETHEWYNISGERTAHISGLRPGNDFIYVLSGLAPGQTQRPISATATEAPEVDNLL 1172
QY 1636 LGLMTYDNGTGLATKSDETGWTTFY----DVDHEGLRINVT-----RPTGVVTSLSHREME 1687
Db 1173 VSDATPDGFR--LSWTADEGVDSFVLKIRDTKKQSEPLEITLLASERTDRITGLREATE 1230
QY 1688 KSTII--DIENSRRDDDVITNL---SSVEASYVVQDVQRNSYQLCNGNGLTVRVMTANGM 1743
Db 1231 YEELYGISGRKSPQVSAIATAMGSPKEITFSDITE-----NSATVSMWVPTAQ 1281
QY 1744 GTSFH-----SEPHVLGATITPTIGRCNISLPMENGLNSTEWRLEQKQKVTIFG 1795
Db 1282 VESFRITYVITGGARSVTVVDTGRTQTRLLRLP-----GVEYLVSVIAVKG-----FE 1331
QY 1796 RKLVRHGRMLLSIDYDRNIRTEKIYDDHRKFTLRITYDQVGRFLLWPSGLAAVNVSYF 1855
Db 1332 ESEPVSGTLTTALDGPGLVTANITDSEAL-----AMQPAIAPVDHYVISY 1378
QY 1856 ENGLAGLORGAMSER-----TDIDKQGRIVSRMFADGKVMWSYLDKSNVLLLSQSORQY 1910
Db 1379 TGDVPEIPTRTVSGNTVEVALINLEPATEYTLRFAEKGPQKSSITFTFTDLDSPRDL 1438
QY 1911 -IFEYDSSDRLLAVTMP--SVARHSMSTHTSIGIYRIINYPPEASNAVIFDYSDDGRIKL 1967
Db 1439 TATEVQSETALLTWPRASVTGTYLLVYESVDGTLKEVVVGPEPETS-----YSLSGLSFS 1493
QY 1968 TSFLGTGRGVFYKYGKLSKLSIYDSTAVTFGY-----DETGVLMK-----2010
Db 1494 THY--TARIQALNGPLRSKMSQTVFTTIGLXPFPFRDCSQAMLNGTTSGLTYIYVNDK 1551
QY 2011 -----VNLQSGGFSCTIR-----YRK-----IGPLVDK-----QIVRFSE 2040
Db 1552 AOKLEVFCDMTSDSGGWIIVFLRRKNGREDFYRNWKAAYAAGFDLKEEFWGLGDALSKITA 1611
QY 2041 EGMVNAFDTYTHDNS-----FRIASIKPVISETPLPVDLYRYDEISGKVEHFGK-FGVITY 2095
Db 1612 QGOYELRVLDLHGEYAYAVYDRFSVGDARTRYKLVGEGYSGTAGDSMAYHNGRSESTFD 1671
QY 2096 YDINQIITAVMTLSKHF-----DTHGRKEVOYEMFRSLMYMTVQYDSMGRIKVR 2147
Db 1672 KDTDSAITNCALSYKGAFFYKNCRVNLMGRYCDNSHSGQVNFHMKGKEYS-----IQFA 1727
QY 2148 ELKGLP 2153
Db 1728 EMKLRP 1733

RESULT 4
TENX HUMAN
ID TENX HUMAN STANDARD; PRT: 4289 AA.
AC P22105; P78530; P78531; Q08424; Q0UMG7;
DT 01-AUG-1991 (Rel. 19, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Tenascin-X precursor (TN-X) (Hexabrachion-like).
GN TNXB OR TNX OR XB OR HXBL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Rowen L., Dankers C., Baskin D., Faust J., Loretz C., Ahearn M.E.,
RA Banca A., Schwartzell S., Smith T.M., Spies T., Hood L.;
RT "Sequence determination of 300 kilobases of the human class III MHC
RT locus.";
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
```

```
RP SEQUENCE OF 1-747 AND 1687-1944 FROM N.A.
RC TISSUE=Leukocyte;
RX MEDLINE=93300909; PubMed=7686164;
RA Bristow J., Tee M.K., Gitelman S.E., Mellon S.H., Miller W.L.;
RT "Tenascin-X: a novel extracellular matrix protein encoded by the human
RT XB gene overlapping P450c21b.";
RL J. Cell Biol. 122:265-278(1993).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM XB-SHORT).
RC TISSUE=Adrenal gland;
RX MEDLINE=96015044; PubMed=8530023;
RA Tee M.K., Thomson A.A., Bristow J., Miller W.L.;
RT "Sequences promoting the transcription of the human XA gene
RT overlapping P450c21A correctly predict the presence of a novel,
RT adrenal-specific, truncated form of tenascin-X.";
RL Genomics 28:171-178(1995).
RN [4]
RP SEQUENCE OF 1-23 FROM N.A.
RC TISSUE=Fetal adrenal gland;
RX MEDLINE=97081760; PubMed=8923003;
RA Speck M., Barry F., Miller W.L.;
RT "Alternate promoters and alternate splicing of human tenascin-X, a
RT gene with 5' and 3' ends buried in other genes.";
RL Hum. Mol. Genet. 5:1749-1758(1996).
RN [5]
RP SEQUENCE OF 3470-4289 FROM N.A.
RX MEDLINE=89367293; PubMed=2475872;
RA Morel Y., Bristow J., Gitelman S.E., Miller W.L.;
RT "Transcript encoded on the opposite strand of the human steroid 21-
RT hydroxylase/complement component C4 gene locus";
RL Proc. Natl. Acad. Sci. U.S.A. 86:6582-6586(1989).
CC -!- FUNCTION: APPEARS TO MEDIATE INTERACTIONS BETWEEN CELLS AND THE
CC EXTRACELLULAR MATRIX. SUBSTRATE-ADHESION MOLECULE THAT APPEARS TO
CC INHIBIT CELL MIGRATION. MAY PLAY A ROLE IN SUPPORTING THE GROWTH
CC OF EPITHELIAL TUMORS.
CC -!- SUBCELLULAR LOCATION: Extracellular matrix.
CC -!- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS; XB (SHOWN HERE) AND
CC XB-SHORT; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN FETAL ADRENAL, IN FETAL
CC TESTIS, FETAL SMOOTH, STRIATED AND CARDIAC MUSCLE. XB-SHORT IS
CC ONLY EXPRESSED IN THE ADRENAL GLAND.
CC -!- DISEASE: ASSOCIATION WITH CONGENITAL ADRENAL HYPERPLASIA.
CC -!- DISEASE: Defects in TNXB are a cause of autosomal recessive
CC Ehlers-Danlos syndrome.
CC -!- SIMILARITY: CONTAINS 19 EGF-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 32 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS A DOMAIN RELATED TO THE C-TERMINAL PART OF
CC THE BETA AND GAMMA CHAINS OF FIBRINOGEN.
CC -!- CAUTION: THERE ARE TWO GENES FOR TN-X: TNXA AND TNXB. TNXA IS
CC A PARTIAL GENE WHICH CAN SOMETIME RECOMBINE WITH TNXB.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U89337; AAB47488.1; -
DR EMBL; AF019413; AAB67981.1; -
DR EMBL; X71923; CAA50739.1; -
DR EMBL; Y13782; CAA74109.1; -
DR EMBL; Y13783; CAA74110.1; -
DR EMBL; U24488; AAB41287.1; -
DR EMBL; U52696; AAC50889.1; -
DR EMBL; M25813; AAA35884.1; -
DR HSSP; P02671; IFDZ.
DR MIM; 600985; -
DR MIM; 606408; -
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR003961; FN.III.
DR InterPro; IPR002181; Fibrinogen_C.
```


homolog of Drosophila Notch.";
Genomics 15:259-264(1993).
[2]
SEQUENCE OF 1551-2170 FROM N.A.
TISSUE=Embryo;
MEDLINE=33048835; Pubmed=1425352;
Franco del Amo F., Smith D.E., Swiatek P.J., Gendron-Maguire M.,
Greenspan R.J., McMahon A.P., Gridley T.;
"Expression pattern of Notch, a mouse homolog of Drosophila Notch,
suggests an important role in early postimplantation mouse
development.";
RL development 115:737-744(1992).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED ALMOST UNIFORMLY IN EARLY EMBRYOS.
CC -1- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC -1- SIMILARITY: CONTAINS 5 ANK REPEATS.
CC -1- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial/
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

EMBL: Z11886; CAA77941.1; .
DR HSP; P00740; IEDM.
DR MGd; MG1:97363; Notchl.
DR InterPro: IPR002110; ANK.
DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR000561; Asx-like.
DR InterPro: IPR000742; EGF-2.
DR InterPro: IPR001881; EGF-Ca.
DR InterPro: IPR001438; EGF-II.
DR InterPro: IPR000800; Notch.
DR Pfam; PF00223; ank; 6.
DR Pfam; PF00008; EGF; 35.
DR Pfam; PF00066; notch; 3.
DR PRINTS; PR00010; EGFBLOOD.
DR PRINTS; PR01452; NOTCH.
DR SMART; SM00248; ANK; 3.
DR SMART; SM00179; EGF_CA; 23.
DR SMART; SM00001; EGF_Like; 11.
DR SMART; SM00004; NL; 2.
DR PROSITE; PS50088; ANK_REPEAT; 2.
DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
DR PROSITE; PS00010; ASX_HYDROXYL; 22.
DR PROSITE; PS00032; EGF_1; 34.
DR PROSITE; PS01186; EGF_2; 27.
DR PROSITE; PS01187; EGF_CA; 21.
KW Differentiation; Neurogenesis; Repeat; ANK repeat; EGF-like domain;
KW Transmembrane; Signal; Glycoprotein.
FT SIGNAL 1 18 POTENTIAL.
FT CHAIN 19 2531 NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1.
FT DOMAIN 19 1725 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1726 1746 POTENTIAL.
FT DOMAIN 1747 2531 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 20 58 EGF-Like 1.
FT DOMAIN 59 99 EGF-Like 2.
FT DOMAIN 102 139 EGF-Like 3.
FT DOMAIN 140 176 EGF-Like 4.
FT DOMAIN 178 215 EGF-Like 5.
FT DOMAIN 218 255 EGF-Like 6.
FT DOMAIN 257 293 EGF-Like 7.
FT DOMAIN 295 333 EGF-Like 8.
FT DOMAIN 335 371 EGF-Like 9.
FT DOMAIN 372 410 EGF-Like 10.
FT DOMAIN 412 450 EGF-Like 11.
FT DOMAIN 452 488 EGF-Like 12.
FT DOMAIN 490 526 EGF-Like 13.
FT DOMAIN 528 564 EGF-Like 14.
FT DOMAIN 564 598 CALCIUM-BINDING (POTENTIAL).

FT	DISULFID	575	589	BY SIMILARITY.	AC	Q07008;
FT	DISULFID	591	600	BY SIMILARITY.	DT	01-NOV-1995 (Rel. 32, Created)
FT	DISULFID	607	618	BY SIMILARITY.	DT	15-JUL-1999 (Rel. 38, Last sequence update)
FT	DISULFID	612	627	BY SIMILARITY.	DT	16-OCT-2001 (Rel. 40, Last annotation update)
FT	DISULFID	629	638	BY SIMILARITY.	DE	Neurogenic locus notch homolog protein 1 precursor.
FT	DISULFID	645	655	BY SIMILARITY.	GN	NOTCH1.
FT	DISULFID	650	664	BY SIMILARITY.	OS	Rattus norvegicus (Rat).
FT	DISULFID	666	675	BY SIMILARITY.	OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
FT	DISULFID	682	693	BY SIMILARITY.	OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
FT	DISULFID	687	702	BY SIMILARITY.	OC	NCBI_TaxID=10116;
FT	DISULFID	704	713	BY SIMILARITY.	RN	[1]
FT	DISULFID	720	730	BY SIMILARITY.	RP	SEQUENCE FROM N.A.
FT	DISULFID	725	739	BY SIMILARITY.	RC	TISSUE-Schwann cell;
FT	DISULFID	741	750	BY SIMILARITY.	RX	MEDLINE=92111383; PubMed=1764995;
FT	DISULFID	757	768	BY SIMILARITY.	RA	Weinmaster G., Roberts V.J., Lemke G.;
FT	DISULFID	762	777	BY SIMILARITY.	RT	"A homolog of Drosophila Notch expressed during mammalian
FT	DISULFID	779	788	BY SIMILARITY.	RT	development";
FT	DISULFID	795	806	BY SIMILARITY.	RL	Development 113:199-205(1991).
FT	DISULFID	800	815	BY SIMILARITY.	RN	[2]
FT	DISULFID	817	826	BY SIMILARITY.	RP	REVISIONS TO 1652-1653.
FT	DISULFID	833	844	BY SIMILARITY.	RA	Weinmaster G.;
FT	DISULFID	838	855	BY SIMILARITY.	RL	Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
FT	DISULFID	857	866	BY SIMILARITY.	CC	-1- FUNCTION: REQUIRED FOR THE CORRECT DIFFERENTIATION OF A NUMBER
FT	DISULFID	873	884	BY SIMILARITY.	CC	OF TISSUES.
FT	DISULFID	878	893	BY SIMILARITY.	CC	-1- SUBCELLULAR LOCATION: Type I membrane protein.
FT	DISULFID	895	904	BY SIMILARITY.	CC	-1- DEVELOPMENTAL STAGE: IN THE EMBRYO, HIGHEST LEVELS OCCUR BETWEEN
FT	DISULFID	911	922	BY SIMILARITY.	CC	DAYS 12 AND 14 AND DECREASE RAPIDLY TO MUCH LOWER LEVELS IN THE
FT	DISULFID	931	931	BY SIMILARITY.	CC	ADULT.
FT	DISULFID	933	942	BY SIMILARITY.	CC	-1- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
FT	DISULFID	987	998	BY SIMILARITY.	CC	-1- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
FT	DISULFID	992	1007	BY SIMILARITY.	CC	-1- SIMILARITY: CONTAINS 5 ANK REPEATS.
FT	DISULFID	1009	1018	BY SIMILARITY.	CC	-----
FT	DISULFID	1025	1036	BY SIMILARITY.	CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
FT	DISULFID	1030	1045	BY SIMILARITY.	CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
FT	DISULFID	1047	1056	BY SIMILARITY.	CC	the European Bioinformatics Institute. There are no restrictions on its
FT	DISULFID	1063	1074	BY SIMILARITY.	CC	use by non-profit institutions as long as its content is in no way
FT	DISULFID				CC	modified and this statement is not removed. Usage by and for commercial
FT	DISULFID				CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
FT	DISULFID				CC	or send an email to license@isb-sib.ch).
FT	DISULFID				CC	-----
Qy	532	VODCPRN--CHNGECVSGV----	CHCPFGFLGADG-----AKAACPVLCSGN--GOYSKG	579	DR	EMBL; X57405; GAA40667.1; -.
Db	453	VNECLSNPCQNDATCLDQIGEFQCICMPYEGYVEINTDECASSPCLHNGHCHMDKIHFEF	512	DR	DR	HSSP; P00740; LIXA.
Qy	580	TCQCYSGWKGAECDDPMNQCIDPSCGGHSGCIDG-----NCVCSAGYKGEHCE-EVDCCLDP	634	DR	DR	InterPro; IPR000152; Asx_hydroxyl.
Qy	513	QCQCPKGFNGHLQYDVDECASTPCKNGAKCLDGPNTVTCVCTGCTHCEVDIDCDP	572	DR	DR	InterPro; IPR000561; EGF-like.
Qy	635	TCSSHGVCVNG-----ECLCSPGNGLNCELARVQCPCQCSGHGYLPD---TGLCSCDPN	687	DR	DR	InterPro; IPR000742; EGF_2.
Db	573	DPCHYGSKDGVATFTCLCQPGYTGHHCEETINECHSQPCRRHGGTCQDRDNSYLCCLKG	632	DR	DR	InterPro; IPR001881; EGF_Ca.
Qy	688	WMGPDCSVEVCSVDGTH-----GVCI-----GCACRCEGWTGAACDQV-----CH--	730	DR	DR	InterPro; IPR001438; EGF_II.
Db	633	TTGPNCEINL--DDCASNPDSGCTCLDKIDGYECACEPGYTGSMCNVNDICAGSPCHNG	690	DR	DR	InterPro; IPR002049; Laminin_EGF.
Qy	731	-----PRCIE-----HCTCKDG-----KCECREGNGRHC	755	DR	DR	Pfam; PF00023; ank; 6.
Db	691	GTCEDGIAGFTCRCPGEGYHDTCLISEVNECSNPNHICACRDGLNGYKCDCAPGHSGTNC	750	DR	DR	Pfam; PF00008; EGF; 36.
Qy	756	TIGROTATETDGGCPDLGNGRCTLGQNSWCQCQCTGWRGPGCNVAMETSCADNKD-NE	814	DR	DR	PRINTS; PR00010; EGFBLLOOD.
Db	751	DIN-----NNECESNP--CVNGGTCKDMTSGYVCTCREGFGSPGNCQTN-NECASNPCLNQ	803	DR	DR	PRINTS; PR01452; NOTCH.
Qy	815	GDGLVDCLDPPC-----CLQSAQNSLLCRGSRD	843	DR	DR	SMART; SM00248; ANK; 5.
Db	804	GTCIDDVAGYKCNCLPYTGATCEVVLAPCATSPCKNSGVCKESD	849	DR	DR	SMART; SM00179; EGF_Ca; 25.
RESULT	6				DR	SMART; SM00001; EGF_like; 10.
NTCL_RAT					DR	SMART; SM00004; NL; 2.
ID					DR	PROSITE; PS50088; ANK_REPEAT; 4.
					DR	PROSITE; PS50297; ANK_REPEAT_REGION; 1.
					DR	PROSITE; PS00010; ASX_HYDROXYL; 22.
					DR	PROSITE; PS00022; EGF_1; 35.
					DR	PROSITE; PS01186; EGF_2; 26.
					DR	PROSITE; PS01187; EGF_CA; 21.
					KW	Differentiation; Neurogenesis; Repeat; ANK repeat; EGF-like domain;
					KW	Transmembrane; Signal; Glycoprotein.
					FT	SIGNAL
					FT	CHAIN
					FT	DOMAIN
					FT	TRANSMEM

1 18
19 2531
19 1723
1724 1746

POTENTIAL.
NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1.
EXTRACELLULAR (POTENTIAL).
POTENTIAL.

Oy 731 -----PRCIE-----RGCTCKDG-----KCECREGNGEHC 755
 Db 691 GTCEDGAGFTCRCEGVDHPTCLSEVNECNSNPCIHGACRDGLNGYKCDAPGWSGTNC 750
 Oy 756 TIGRQTAGTETDGGCDLCLNGRCTLGONSQVCQGTGWRGPGCNVAMEYSCAOKD-NE 814
 Db 751 DIN-----NNECESNP--CVNGGTCKDMTSGYVCVTCRGGFGSPNCQTNI-NECASNPCLNQ 803
 Oy 815 GDGLVDCLDLDPDC-----CLQSACQNSLLCRGSRD 843
 Db 804 GTCIDDVAGYKCNCPLPYTGATCEVWLAPCATSPCKNSGVCKESED 849
 RESULT 7
 NTC3_MOUSE
 ID NTC3_MOUSE STANDARD; PRT; 2318 AA.
 Q61982:
 01-NOV-1997 (Rel. 35, Created)
 01-NOV-1997 (Rel. 35, Last sequence update)
 16-OCT-2001 (Rel. 40, Last annotation update)
 Neurogenic locus notch 3 protein.
 GN NOTCH3.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RC STRAIN=ICR X SWISS WEBSTER;
 RX MEDLINE=95001556; PubMed=7918097;
 RA Lardelli M., Dalstrand J., Lendahl U.;
 RT "The novel Notch homologue mouse Notch 3 lacks specific epidermal
 growth factor-repeats and is expressed in proliferating
 neuroepithelium";
 RL Mech. Dev. 46:123-136(1994).
 CC -1- FUNCTION: NOTCH 1, 2 AND 3 PLAY A COMBINATIONAL ROLE DURING
 VARIOUS CELL FATE DECISIONS AND MORPHOLOGICAL MOVEMENTS IN THE
 DEVELOPING CNS AND PROBABLY OTHER REGIONS OF THE EMBRYO.
 CC -1- TISSUE SPECIFICITY: PROLIFERATING NEUROEPITHELIUM.
 CC -1- DEVELOPMENTAL STAGE: CNS DEVELOPMENT.
 CC -1- SIMILARITY: CONTAINS 34 EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
 CC -1- SIMILARITY: CONTAINS 5 ANK REPEATS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X74760; CAA52776.1; --
 DR HSSP; P00740; IIXA.
 DR MGD; MGI:99460; Notch3.
 DR InterPro; IPR002110; ANK.
 DR InterPro; IPR000152; Asx_hydroxyl.
 DR InterPro; IPR000561; EGF-like.
 DR InterPro; IPR000742; EGF-2.
 DR InterPro; IPR001881; EGF_Ca.
 DR InterPro; IPR001438; EGF_II.
 DR InterPro; IPR000800; Notch.
 DR Pfam; PF000023; ank; 6.
 DR Pfam; PF00008; EGF; 34.
 DR Pfam; PF00066; notch; 3.
 DR PRINTS; PR00010; EGFBL00D.
 DR PRINTS; PR01452; NOTCH.
 DR SMART; SM00248; ANK; 5.
 DR SMART; SM00179; EGF_Ca; 19.
 DR SMART; SM00001; EGF_like; 13.
 DR SMART; SM00004; NL; 3.
 DR PROSITE; PS50088; ANK_REPEAT; 4.

DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
 DR PROSITE; PS00010; ASX_HYDROXYL; 18.
 DR PROSITE; PS00022; EGF_1; 33.
 DR PROSITE; PS01186; EGF_2; 27.
 DR PROSITE; PS01187; EGF_CA; 17.
 KW Differentiation; Neurogenesis; Repeat; EGF-like domain; Transmembrane;
 KW ANK repeat; Glycoprotein.
 FT DOMAIN 1 1643 EXTRACELLULAR.
 FT TRANSMEM 1644 1664 POTENTIAL.
 FT DOMAIN 1665 2318 CYTOPLASMIC.
 FT DOMAIN 2242 2261 PEST.
 FT DOMAIN 39 78 EGF-LIKE 1.
 FT DOMAIN 79 119 EGF-LIKE 2.
 FT DOMAIN 120 157 EGF-LIKE 3.
 FT DOMAIN 159 196 EGF-LIKE 4.
 FT DOMAIN 198 235 EGF-LIKE 5.
 FT DOMAIN 237 273 EGF-LIKE 6.
 FT DOMAIN 275 313 EGF-LIKE 7.
 FT DOMAIN 315 351 EGF-LIKE 8.
 FT DOMAIN 352 390 EGF-LIKE 9.
 FT DOMAIN 392 430 CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 432 468 EGF-LIKE 11.
 FT DOMAIN 470 506 EGF-LIKE 12.
 FT DOMAIN 508 544 EGF-LIKE 13.
 FT DOMAIN 546 581 EGF-LIKE 14.
 FT DOMAIN 583 619 EGF-LIKE 15.
 FT DOMAIN 621 656 EGF-LIKE 16.
 FT DOMAIN 658 694 EGF-LIKE 17.
 FT DOMAIN 696 731 EGF-LIKE 18.
 FT DOMAIN 735 771 EGF-LIKE 19.
 FT DOMAIN 772 809 EGF-LIKE 20.
 FT DOMAIN 811 848 EGF-LIKE 21.
 FT DOMAIN 850 886 EGF-LIKE 22.
 FT DOMAIN 888 923 EGF-LIKE 23.
 FT DOMAIN 925 961 EGF-LIKE 24.
 FT DOMAIN 963 999 EGF-LIKE 25.
 FT DOMAIN 1001 1035 EGF-LIKE 26.
 FT DOMAIN 1037 1083 EGF-LIKE 27.
 FT DOMAIN 1085 1121 EGF-LIKE 28.
 FT DOMAIN 1123 1159 EGF-LIKE 29.
 FT DOMAIN 1161 1204 EGF-LIKE 30.
 FT DOMAIN 1206 1245 EGF-LIKE 31.
 FT DOMAIN 1247 1288 EGF-LIKE 32.
 FT DOMAIN 1290 1326 EGF-LIKE 33.
 FT DOMAIN 1336 1374 EGF-LIKE 34.
 FT REPEAT 1388 1428 LIN/NOTCH 1.
 FT REPEAT 1429 1467 LIN/NOTCH 2.
 FT REPEAT 1468 1503 LIN/NOTCH 3.
 FT REPEAT 1839 1868 ANK 1.
 FT REPEAT 1872 1902 ANK 2.
 FT REPEAT 1906 1935 ANK 3.
 FT REPEAT 1939 1968 ANK 4.
 FT REPEAT 1972 2001 ANK 5.
 FT DISULFID 43 55 BY SIMILARITY.
 FT DISULFID 49 66 BY SIMILARITY.
 FT DISULFID 68 77 BY SIMILARITY.
 FT DISULFID 83 94 BY SIMILARITY.
 FT DISULFID 88 107 BY SIMILARITY.
 FT DISULFID 109 118 BY SIMILARITY.
 FT DISULFID 124 135 BY SIMILARITY.
 FT DISULFID 129 145 BY SIMILARITY.
 FT DISULFID 147 156 BY SIMILARITY.
 FT DISULFID 163 175 BY SIMILARITY.
 FT DISULFID 169 184 BY SIMILARITY.
 FT DISULFID 186 195 BY SIMILARITY.
 FT DISULFID 202 213 BY SIMILARITY.
 FT DISULFID 207 223 BY SIMILARITY.
 FT DISULFID 225 234 BY SIMILARITY.
 FT DISULFID 241 252 BY SIMILARITY.
 FT DISULFID 246 261 BY SIMILARITY.
 FT DISULFID 263 272 BY SIMILARITY.
 FT DISULFID 279 292 BY SIMILARITY.
 FT DISULFID 286 301 BY SIMILARITY.

FT	DISULFID	303	312	BY SIMILARITY.
FT	DISULFID	319	319	BY SIMILARITY.
FT	DISULFID	324	339	BY SIMILARITY.
FT	DISULFID	341	350	BY SIMILARITY.
FT	DISULFID	356	367	BY SIMILARITY.
FT	DISULFID	361	378	BY SIMILARITY.
FT	DISULFID	380	389	BY SIMILARITY.
FT	DISULFID	396	409	BY SIMILARITY.
FT	DISULFID	403	418	BY SIMILARITY.
FT	DISULFID	420	429	BY SIMILARITY.
FT	DISULFID	436	447	BY SIMILARITY.
FT	DISULFID	441	456	BY SIMILARITY.
FT	DISULFID	458	467	BY SIMILARITY.
FT	DISULFID	474	485	BY SIMILARITY.
FT	DISULFID	479	494	BY SIMILARITY.
FT	DISULFID	496	505	BY SIMILARITY.
FT	DISULFID	512	523	BY SIMILARITY.
FT	DISULFID	517	532	BY SIMILARITY.
FT	DISULFID	534	543	BY SIMILARITY.
FT	DISULFID	550	560	BY SIMILARITY.
FT	DISULFID	555	569	BY SIMILARITY.
FT	DISULFID	571	580	BY SIMILARITY.
FT	DISULFID	587	598	BY SIMILARITY.
FT	DISULFID	592	607	BY SIMILARITY.
FT	DISULFID	609	618	BY SIMILARITY.
FT	DISULFID	625	635	BY SIMILARITY.
FT	DISULFID	630	644	BY SIMILARITY.
FT	DISULFID	646	655	BY SIMILARITY.
FT	DISULFID	662	673	BY SIMILARITY.
FT	DISULFID	667	682	BY SIMILARITY.
FT	DISULFID	684	693	BY SIMILARITY.
FT	DISULFID	700	710	BY SIMILARITY.
FT	DISULFID	705	719	BY SIMILARITY.
FT	DISULFID	721	730	BY SIMILARITY.
FT	DISULFID	739	750	BY SIMILARITY.
FT	DISULFID	744	759	BY SIMILARITY.
FT	DISULFID	761	770	BY SIMILARITY.
FT	DISULFID	776	787	BY SIMILARITY.
FT	DISULFID	781	797	BY SIMILARITY.
FT	DISULFID	799	808	BY SIMILARITY.
FT	DISULFID	815	827	BY SIMILARITY.
FT	DISULFID	831	836	BY SIMILARITY.
FT	DISULFID	838	847	BY SIMILARITY.
FT	DISULFID	854	865	BY SIMILARITY.
FT	DISULFID	859	874	BY SIMILARITY.
FT	DISULFID	876	885	BY SIMILARITY.
FT	DISULFID	892	902	BY SIMILARITY.
FT	DISULFID	897	911	BY SIMILARITY.
FT	DISULFID	913	922	BY SIMILARITY.
FT	DISULFID	929	940	BY SIMILARITY.
FT	DISULFID	934	949	BY SIMILARITY.
FT	DISULFID	951	960	BY SIMILARITY.
FT	DISULFID	967	978	BY SIMILARITY.
FT	DISULFID	972	987	BY SIMILARITY.
FT	DISULFID	989	998	BY SIMILARITY.
FT	DISULFID	1005	1016	BY SIMILARITY.
FT	DISULFID	1010	1023	BY SIMILARITY.
FT	DISULFID	1025	1034	BY SIMILARITY.
FT	DISULFID	1041	1062	BY SIMILARITY.
FT	DISULFID	1056	1071	BY SIMILARITY.
FT	DISULFID	1073	1082	BY SIMILARITY.
FT	DISULFID	1089	1100	BY SIMILARITY.
FT	DISULFID	1094	1109	BY SIMILARITY.
FT	DISULFID	1111	1120	BY SIMILARITY.
FT	DISULFID	1127	1138	BY SIMILARITY.
FT	DISULFID	1132	1147	BY SIMILARITY.
FT	DISULFID	1149	1158	BY SIMILARITY.
FT	DISULFID	1165	1183	BY SIMILARITY.
FT	DISULFID	1177	1192	BY SIMILARITY.
FT	DISULFID	1194	1203	BY SIMILARITY.
Query Match				3.1%; Score 450.5; DB 1; Length 2318;
Best Local Similarity				26.2%; Pred. No. 2.5e-16;
<hr/>				
Matches 126; Conservative 37; Mismatches 136; Indels 181; Gaps 24;				
Qy	531	SVQDC-PRNCHNGECVSGV-----CHCFPGFLGA-----	559	
Db	546	NVDDCSPDPCH-HGRCDGIIASFSCACAPGYTGIRCESQVDECRSQPCRYGKGLDLVDK	604	
Qy	560	-----DCAKAACP--VLCSGNGQYKGTCCQCYSGWKGAECDDPMNQ	598	
Db	605	YLCRCPPGTTGVNCEVNIDDCASNCTFCGCRDGINRYD---CVCQPGFTGLCNVNE	661	
Qy	599	CIDPSCGGHSGCIDG-----	617	NCVC
Db	662	CASSPCGEGGSCVDGNGFHCILCPGSLPPLCLPNHPHCAHKPCSHGVCHDAPGGFRCV	721	
Qy	618	SAGYKGEHCEEV-----DCIDPTCSSHRGV-----VNGECLCSPWGLNCELARVCPDQC	669	
Db	722	EPGWSGPCSOSLAPDACESQPCAGGTCTSDGIGFRCTCAFGHGHQCEVLSPTPSLC	781	
Qy	670	---SGHGTLPD-TGLCSDPNMGPDC--SVEVC--SVDCGTHGVCI-----GGACRCEG	718	
Db	782	EHGGHCESDPDLTVCSPPGQWQGPCQDQVDECAGASPCGPHGTCTNLPGNFRCTCHRG	841	
Qy	719	WTGAACDQRV--CHPR-CIEHGTCKDG---KCECREGHWG-----EHCTIGROTACT	764	
Db	842	YTGPFCDQDIDCDPNPCLHGGSCQDGVGSFSCCLDGFAGRPCARDVDECLSSPCGPGT	901	
Qy	765	ETD-----GCP-----DL-----CNGNGRCTLGQNSWCQVCQTGWRGPGC-	799	
Db	902	CITDHVAFTCACPPGYGVGFHCEIDLPCSPSSCFNGGTCVGVGSFSCLCRPGYGTGHCQ	961	
Qy	800	-----NVAMETSCADNKDNGDGLVDCCLDP-DCCLQSAQNSLLC	838	
Db	962	YEADPCFSRPLUHGIGCNTPHGFECTC-----REGFTGSOQNVPDMCWSQAPCGNGRC	1016	
<hr/>				
RESULT 8				
NTCL_HUMAN				
ID	NTCL_HUMAN	STANDARD;	PRT; 2444 AA.	
AC	P46531;			
DT	01-NOV-1995 (Rel. 32, Created)			
DT	01-NOV-1995 (Rel. 32, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Neurogenic locus notch protein homolog 1 precursor (Translocation-			
DE	associated notch protein TAN-1) (Fragment).			
GN	NOTCH1 OR TAN1.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=91347367; PubMed=1831692;			
RA	Ellisen L.W., Bird J., West D.C., Soreng A.L., Reynolds T.C.,			
RA	Smith S.D., Sklar J.;			
RT	"TAN-1, the human homolog of the Drosophila notch gene, is broken by			
RT	chromosomal translocations in T lymphoblastic neoplasms.";			
RL	Cell 66:649-661(1991).			
CC	-!- FUNCTION: MAY BE IMPORTANT FOR NORMAL LYMPHOCYTE FUNCTION. IN			
CC	ALTERED FORM, MAY CONTRIBUTE TO TRANSFORMATION OR PROGRESSION			
CC	IN SOME T-CELL NEOPLASMS.			
CC	-!- SUBCELLULAR LOCATION: Type I membrane protein.			
CC	-!- TISSUE SPECIFICITY: IN FETAL TISSUES MOST ABUNDANT IN SPLEEN,			
CC	BRAIN STEM AND LUNG. ALSO PRESENT IN MOST ADULT TISSUES WHERE IT			
CC	IS FOUND MAINLY IN LYMPHOID TISSUES.			
CC	-!- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.			
CC	-!- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.			
CC	-!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.			
CC	-!- SIMILARITY: CONTAINS 5 ANK REPEATS.			
<hr/>				
This SWISS-PROT entry is copyright. It is produced through a collaboration				
between the Swiss Institute of Bioinformatics and the EMBL outstation -				
the European Bioinformatics Institute. There are no restrictions on its				
use by non-profit institutions as long as its content is in no way				

CC	modified and this statement is not removed. Usage by and for commercial	FT	REPEAT	1991	ANK 2.
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/	FT	REPEAT	1995	ANK 3.
CC	or send an email to license@isb-sib.ch).	FT	REPEAT	2028	ANK 4.
CC	-----	FT	REPEAT	2061	ANK 5.
DR	EMBL; M73980; AAA60614.1; -.	FT	DOMAIN	1576	POLY-VAL.
DR	HSSP; P00740; IIXA.	FT	DOMAIN	1662	POLY-ARG.
DR	MIM; 190198; -.	FT	DOMAIN	1729	POLY-ARG.
DR	InterPro; IPR002110; ANK.	FT	DOMAIN	1741	POLY-PRO.
DR	InterPro; IPR000152; Asx_hydroxyl.	FT	DOMAIN	1902	POLY-ALA.
DR	InterPro; IPR000561; EGF-like.	FT	DOMAIN	1902	POLY-GLY.
DR	InterPro; IPR000742; EGF.2.	FT	DOMAIN	2260	POLY-GLY.
DR	InterPro; IPR001881; EGF.Ca.	FT	DOMAIN	2404	POLY-GLN.
DR	InterPro; IPR000800; Notch.	FT	DOMAIN	2411	POLY-PRO.
DR	Pfam; PF00023; ank; 6.	FT	DISULFID	24	BY SIMILARITY
DR	Pfam; PF00008; EGF; 36.	FT	DISULFID	31	BY SIMILARITY
DR	Pfam; PF00066; notch; 3.	FT	DISULFID	48	BY SIMILARITY
DR	SMART; SM00248; ANK; 5.	FT	DISULFID	63	BY SIMILARITY
DR	SMART; SM00179; EGF.CA; 22.	FT	DISULFID	68	BY SIMILARITY
DR	SMART; SM00001; EGF-like; 13.	FT	DISULFID	89	BY SIMILARITY
DR	SMART; SM00004; NL; 2.	FT	DISULFID	89	BY SIMILARITY
DR	PROSITE; PS50088; ANK_REPEAT; 4.	FT	DISULFID	106	BY SIMILARITY
DR	PROSITE; PS50297; ANK_REP_REGION; 1.	FT	DISULFID	111	BY SIMILARITY
DR	PROSITE; PS00010; ASX_HYDROXYL; 20.	FT	DISULFID	129	BY SIMILARITY
DR	PROSITE; PS00022; EGF_1; 34.	FT	DISULFID	144	BY SIMILARITY
DR	PROSITE; PS01186; EGF_2; 26.	FT	DISULFID	149	BY SIMILARITY
DR	PROSITE; PS01187; EGF.CA; 18.	FT	DISULFID	166	BY SIMILARITY
KW	Differentiation; Neurogenesis; Repeat; ANK repeat; EGF-like domain;	FT	DISULFID	182	BY SIMILARITY
KW	Transmembrane; Signal; Glycoprotein.	FT	DISULFID	189	BY SIMILARITY
FT	SIGNAL	FT	DISULFID	206	BY SIMILARITY
FT	CHAIN	FT	DISULFID	222	BY SIMILARITY
FT	DOMAIN	FT	DISULFID	227	BY SIMILARITY
FT	TRANSSEM	FT	DISULFID	245	BY SIMILARITY
FT	DOMAIN	FT	DISULFID	261	BY SIMILARITY
FT	DOMAIN	FT	DISULFID	266	BY SIMILARITY
FT	DOMAIN	FT	DISULFID	283	BY SIMILARITY
FT	DOMAIN	FT	DISULFID	292	BY SIMILARITY
FT	DOMAIN	FT	DISULFID	299	BY SIMILARITY
FT	DOMAIN	FT	DISULFID	306	BY SIMILARITY
FT	DOMAIN	FT	DISULFID	321	BY SIMILARITY
FT	DOMAIN	FT	DISULFID	323	BY SIMILARITY
FT	DOMAIN	FT	DISULFID	339	BY SIMILARITY
FT	DOMAIN	FT	DISULFID	344	BY SIMILARITY
FT	DOMAIN	FT	DISULFID	361	BY SIMILARITY
FT	DOMAIN	FT	DISULFID	370	BY SIMILARITY
FT	DOMAIN	FT	DISULFID	376	BY SIMILARITY
FT	DOMAIN	FT	DISULFID	381	BY SIMILARITY
FT	DOMAIN	FT	DISULFID	400	BY SIMILARITY
FT	DOMAIN	FT	DISULFID	416	BY SIMILARITY
FT	DOMAIN	FT	DISULFID	423	BY SIMILARITY
FT	DOMAIN	FT	DISULFID	440	BY SIMILARITY
FT	DOMAIN	FT	DISULFID	449	BY SIMILARITY
FT	DOMAIN	FT	DISULFID	456	BY SIMILARITY
FT	DOMAIN	FT	DISULFID	461	BY SIMILARITY
FT	DOMAIN	FT	DISULFID	478	BY SIMILARITY
FT	DOMAIN	FT	DISULFID	487	BY SIMILARITY
FT	DOMAIN	FT	DISULFID	494	BY SIMILARITY
FT	DOMAIN	FT	DISULFID	499	BY SIMILARITY
FT	DOMAIN	FT	DISULFID	516	BY SIMILARITY
FT	DOMAIN	FT	DISULFID	525	BY SIMILARITY
FT	DOMAIN	FT	DISULFID	532	BY SIMILARITY
FT	DOMAIN	FT	DISULFID	537	BY SIMILARITY
FT	DOMAIN	FT	DISULFID	552	BY SIMILARITY
FT	DOMAIN	FT	DISULFID	553	BY SIMILARITY
FT	DOMAIN	FT	DISULFID	554	BY SIMILARITY
FT	DOMAIN	FT	DISULFID	570	BY SIMILARITY
FT	DOMAIN	FT	DISULFID	575	BY SIMILARITY
FT	DOMAIN	FT	DISULFID	589	BY SIMILARITY
FT	DOMAIN	FT	DISULFID	591	BY SIMILARITY
FT	DOMAIN	FT	DISULFID	600	BY SIMILARITY
FT	DOMAIN	FT	DISULFID	607	BY SIMILARITY
FT	DOMAIN	FT	DISULFID	612	BY SIMILARITY
FT	DOMAIN	FT	DISULFID	627	BY SIMILARITY
FT	DOMAIN	FT	DISULFID	629	BY SIMILARITY
FT	DOMAIN	FT	DISULFID	645	BY SIMILARITY
FT	DOMAIN	FT	DISULFID	650	BY SIMILARITY
FT	DOMAIN	FT	DISULFID	666	

AC P21783;
 DT 01-MAY-1991 (Rel. 18, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Neurogenic locus notch protein homolog precursor (NOTCH protein).
 GN NOTCH.
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
 OC Xenopodinae; Xenopus.
 OX NCBI_TaxID=8355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90385285; PubMed=2402639;
 RA Coffman C., Harris W., Kintner C.;
 RT "Notch, the Xenopus homolog of Drosophila notch.";
 RL Science 249:1438-1441(1990).
 RN [2]
 RP REVISIONS TO 1759-1782.
 RA Kintner C.;
 RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED ALMOST UNIFORMLY IN EARLY EMBRYOS.
 CC -1- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
 CC -1- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
 CC -1- SIMILARITY: CONTAINS 6 ANK REPEATS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M33874; AB020339.1; -;
 DR PIR; A35844; A35844.
 DR HSP; P00740; IEDM.
 DR InterPro; IPR002110; ANK.
 DR InterPro; IPR000152; Asx_hydroxyl.
 DR InterPro; IPR000561; EGF-like.
 DR InterPro; IPR000742; EGF_2.
 DR InterPro; IPR001881; EGF_Ca.
 DR InterPro; IPR001438; EGF-II.
 DR InterPro; IPR000800; Notch.
 DR Pfam; PF00023; ank; 6.
 DR Pfam; PF00008; EGF; 36.
 DR Pfam; PF00066; notch; 3.
 DR PRINTS; PR00010; EGFBLD.
 DR PRINTS; PR01452; NOTCH.
 DR SMART; SM00248; ANK; 5.
 DR SMART; SM00179; EGF_CA; 23.
 DR SMART; SM00001; EGF-like; 11.
 DR SMART; SM00004; NL; 2.
 DR PROSITE; PS50088; ANK_REPEAT; 4.
 DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
 DR PROSITE; PS00010; ASX_HYDROXYL; 23.
 DR PROSITE; PS00022; EGF_1; 34.
 DR PROSITE; PS01186; EGF_2; 29.
 DR PROSITE; PS01187; EGF_CA; 21.
 KW Differentiation; Neurogenesis; Repeat; ANK repeat; EGF-like domain;
 KW Transmembrane; Signal; Glycoprotein.
 FT SIGNAL 1 19 POTENTIAL.
 FT CHAIN 20 2524 NEUROGENIC LOCUS NOTCH PROTEIN HOMOLOG.
 FT DOMAIN 20 1728 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1729 1750 POTENTIAL.
 FT DOMAIN 1751 2524 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 20 57 EGF-LIKE 1.
 FT DOMAIN 58 99 EGF-LIKE 2.
 FT DOMAIN 102 140 EGF-LIKE 3.
 FT DOMAIN 141 177 EGF-LIKE 4.
 FT DOMAIN 179 215 EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).

217 DOMAIN
 254 FT
 292 FT
 332 FT
 370 FT
 409 FT
 449 FT
 487 FT
 525 FT
 563 FT
 600 FT
 638 FT
 675 FT
 713 FT
 750 FT
 788 FT
 826 FT
 866 FT
 904 FT
 942 FT
 980 FT
 1018 FT
 1056 FT
 1094 FT
 1142 FT
 1180 FT
 1218 FT
 1256 FT
 1294 FT
 1332 FT
 1370 FT
 1408 FT
 1446 FT
 1484 FT
 1522 FT
 1560 FT
 1598 FT
 1636 FT
 1674 FT
 1712 FT
 1750 FT
 1788 FT
 1826 FT
 1864 FT
 1902 FT
 1940 FT
 1978 FT
 2016 FT
 2054 FT
 2092 FT
 2130 FT
 2168 FT
 2206 FT
 2244 FT
 2282 FT
 2320 FT
 2358 FT
 2396 FT
 2434 FT
 2472 FT
 2510 FT
 2548 FT
 2586 FT
 2624 FT
 2662 FT
 2700 FT
 2738 FT
 2776 FT
 2814 FT
 2852 FT
 2890 FT
 2928 FT
 2966 FT
 3004 FT
 3042 FT
 3080 FT
 3118 FT
 3156 FT
 3194 FT
 3232 FT
 3270 FT
 3308 FT
 3346 FT
 3384 FT
 3422 FT
 3460 FT
 3498 FT
 3536 FT
 3574 FT
 3612 FT
 3650 FT
 3688 FT
 3726 FT
 3764 FT
 3802 FT
 3840 FT
 3878 FT
 3916 FT
 3954 FT
 3992 FT
 4030 FT
 4068 FT
 4106 FT
 4144 FT
 4182 FT
 4220 FT
 4258 FT
 4296 FT
 4334 FT
 4372 FT
 4410 FT
 4448 FT
 4486 FT
 4524 FT
 4562 FT
 4600 FT
 4638 FT
 4676 FT
 4714 FT
 4752 FT
 4790 FT
 4828 FT
 4866 FT
 4904 FT
 4942 FT
 4980 FT
 5018 FT
 5056 FT
 5094 FT
 5132 FT
 5170 FT
 5208 FT
 5246 FT
 5284 FT
 5322 FT
 5360 FT
 5398 FT
 5436 FT
 5474 FT
 5512 FT
 5550 FT
 5588 FT
 5626 FT
 5664 FT
 5702 FT
 5740 FT
 5778 FT
 5816 FT
 5854 FT
 5892 FT
 5930 FT
 5968 FT
 6006 FT
 6044 FT
 6082 FT
 6120 FT
 6158 FT
 6196 FT
 6234 FT
 6272 FT
 6310 FT
 6348 FT
 6386 FT
 6424 FT
 6462 FT
 6500 FT
 6538 FT
 6576 FT
 6614 FT
 6652 FT
 6690 FT
 6728 FT
 6766 FT
 6804 FT
 6842 FT
 6880 FT
 6918 FT
 6956 FT
 6994 FT
 7032 FT
 7070 FT
 7108 FT
 7146 FT
 7184 FT
 7222 FT
 7260 FT
 7298 FT
 7336 FT
 7374 FT
 7412 FT
 7450 FT
 7488 FT
 7526 FT
 7564 FT
 7602 FT
 7640 FT
 7678 FT
 7716 FT
 7754 FT
 7792 FT
 7830 FT
 7868 FT
 7906 FT
 7944 FT
 7982 FT
 8020 FT
 8058 FT
 8096 FT
 8134 FT
 8172 FT
 8210 FT
 8248 FT
 8286 FT
 8324 FT
 8362 FT
 8400 FT
 8438 FT
 8476 FT
 8514 FT
 8552 FT
 8590 FT
 8628 FT
 8666 FT
 8704 FT
 8742 FT
 8780 FT
 8818 FT
 8856 FT
 8894 FT
 8932 FT
 8970 FT
 9008 FT
 9046 FT
 9084 FT
 9122 FT
 9160 FT
 9198 FT
 9236 FT
 9274 FT
 9312 FT
 9350 FT
 9388 FT
 9426 FT
 9464 FT
 9502 FT
 9540 FT
 9578 FT
 9616 FT
 9654 FT
 9692 FT
 9730 FT
 9768 FT
 9806 FT
 9844 FT
 9882 FT
 9920 FT
 9958 FT
 9996 FT
 10034 FT
 10072 FT
 10110 FT
 10148 FT
 10186 FT
 10224 FT
 10262 FT
 10300 FT
 10338 FT
 10376 FT
 10414 FT
 10452 FT
 10490 FT
 10528 FT
 10566 FT
 10604 FT
 10642 FT
 10680 FT
 10718 FT
 10756 FT
 10794 FT
 10832 FT
 10870 FT
 10908 FT
 10946 FT
 10984 FT
 11022 FT
 11060 FT
 11098 FT
 11136 FT
 11174 FT
 11212 FT
 11250 FT
 11288 FT
 11326 FT
 11364 FT
 11402 FT
 11440 FT
 11478 FT
 11516 FT
 11554 FT
 11592 FT
 11630 FT
 11668 FT
 11706 FT
 11744 FT
 11782 FT
 11820 FT
 11858 FT
 11896 FT
 11934 FT
 11972 FT
 12010 FT
 12048 FT
 12086 FT
 12124 FT
 12162 FT
 12200 FT
 12238 FT
 12276 FT
 12314 FT
 12352 FT
 12390 FT
 12428 FT
 12466 FT
 12504 FT
 12542 FT
 12580 FT
 12618 FT
 12656 FT
 12694 FT
 12732 FT
 12770 FT
 12808 FT
 12846 FT
 12884 FT
 12922 FT
 12960 FT
 12998 FT
 13036 FT
 13074 FT
 13112 FT
 13150 FT
 13188 FT
 13226 FT
 13264 FT
 13302 FT
 13340 FT
 13378 FT
 13416 FT
 13454 FT
 13492 FT
 13530 FT
 13568 FT
 13606 FT
 13644 FT
 13682 FT
 13720 FT
 13758 FT
 13796 FT
 13834 FT
 13872 FT
 13910 FT
 13948 FT
 13986 FT
 14024 FT
 14062 FT
 14100 FT
 14138 FT
 14176 FT
 14214 FT
 14252 FT
 14290 FT
 14328 FT
 14366 FT
 14404 FT
 14442 FT
 14480 FT
 14518 FT
 14556 FT
 14594 FT
 14632 FT
 14670 FT
 14708 FT
 14746 FT
 14784 FT
 14822 FT
 14860 FT
 14898 FT
 14936 FT
 14974 FT
 15012 FT
 15050 FT
 15088 FT
 15126 FT
 15164 FT
 15202 FT
 15240 FT
 15278 FT
 15316 FT
 15354 FT
 15392 FT
 15430 FT
 15468 FT
 15506 FT
 15544 FT
 15582 FT
 15620 FT
 15658 FT
 15696 FT
 15734 FT
 15772 FT
 15810 FT
 15848 FT
 15886 FT
 15924 FT
 15962 FT
 16000 FT
 16038 FT
 16076 FT
 16114 FT
 16152 FT
 16190 FT
 16228 FT
 16266 FT
 16304 FT
 16342 FT
 16380 FT
 16418 FT
 16456 FT
 16494 FT
 16532 FT
 16570 FT
 16608 FT
 16646 FT
 16684 FT
 16722 FT
 16760 FT
 16798 FT
 16836 FT
 16874 FT
 16912 FT
 16950 FT
 16988 FT
 17026 FT
 17064 FT
 17102 FT
 17140 FT
 17178 FT
 17216 FT
 17254 FT
 17292 FT
 17330 FT
 17368 FT
 17406 FT
 17444 FT
 17482 FT
 17520 FT
 17558 FT
 17596 FT
 17634 FT
 17672 FT
 17710 FT
 17748 FT
 17786 FT
 17824 FT
 17862 FT
 17900 FT
 17938 FT
 17976 FT
 18014 FT
 18052 FT
 18090 FT
 18128 FT
 18166 FT
 18204 FT
 18242 FT
 18280 FT
 18318 FT
 18356 FT
 18394 FT
 18432 FT
 18470 FT
 18508 FT
 18546 FT
 18584 FT
 18622 FT
 18660 FT
 18698 FT
 18736 FT
 18774 FT
 18812 FT
 18850 FT
 18888 FT
 18926 FT
 18964 FT
 19002 FT
 19040 FT
 19078 FT
 19116 FT
 19154 FT
 19192 FT
 19230 FT
 19268 FT
 19306 FT
 19344 FT
 19382 FT
 19420 FT
 19458 FT
 19496 FT
 19534 FT
 19572 FT
 19610 FT
 19648 FT
 19686 FT
 19724 FT
 19762 FT
 19800 FT
 19838 FT
 19876 FT
 19914 FT
 19952 FT
 19990 FT
 20028 FT
 20066 FT
 20104 FT
 20142 FT
 20180 FT
 20218 FT
 20256 FT
 20294 FT
 20332 FT
 20370 FT
 20408 FT
 20446 FT
 20484 FT
 20522 FT
 20560 FT
 20598 FT
 20636 FT
 20674 FT
 20712 FT
 20750 FT
 20788 FT
 20826 FT
 20864 FT
 20902 FT
 20940 FT
 20978 FT
 21016 FT
 21054 FT
 21092 FT
 21130 FT
 21168 FT
 21206 FT
 21244 FT
 21282 FT
 21320 FT
 21358 FT
 21396 FT
 21434 FT
 21472 FT
 21510 FT
 21548 FT
 21586 FT
 21624 FT
 21662 FT
 21700 FT
 21738 FT
 21776 FT
 21814 FT
 21852 FT
 21890 FT
 21928 FT
 21966 FT
 22004 FT
 22042 FT
 22080 FT
 22118 FT
 22156 FT
 22194 FT
 22232 FT
 22270 FT
 22308 FT
 22346 FT
 22384 FT
 22422 FT
 22460 FT
 22498 FT
 22536 FT
 22574 FT
 22612 FT
 22650 FT
 22688 FT
 22726 FT
 22764 FT
 22802 FT
 22840 FT
 22878 FT
 22916 FT
 22954 FT
 22992 FT
 23030 FT
 23068 FT
 23106 FT
 23144 FT
 23182 FT
 23220 FT
 23258 FT
 23296 FT
 23334 FT
 23372 FT
 23410 FT
 23448 FT
 23486 FT
 23524 FT
 23562 FT
 23600 FT
 23638 FT
 23676 FT
 23714 FT
 23752 FT
 23790 FT
 23828 FT
 23866 FT
 23904 FT
 23942 FT
 23980 FT
 24018 FT
 24056 FT
 24094 FT
 24132 FT
 24170 FT
 24208 FT
 24246 FT
 24284 FT
 24322 FT
 24360 FT
 24398 FT
 24436 FT
 24474 FT
 24512 FT
 24550 FT
 24588 FT
 24626 FT
 24664 FT
 24702 FT
 24740 FT
 24778 FT
 24816 FT
 24854 FT
 24892 FT
 24930 FT
 24968 FT
 25006 FT
 25044 FT
 25082 FT
 25120 FT
 25158 FT
 25196 FT
 25234 FT
 25272 FT
 25310 FT
 25348 FT
 25386 FT
 25424 FT
 25462 FT
 25500 FT
 25538 FT
 25576 FT
 25614 FT
 25652 FT
 25690 FT
 25728 FT
 25766 FT
 25804 FT
 25842 FT
 25880 FT
 25918 FT
 25956 FT
 25994 FT
 26032 FT
 26070 FT
 26108 FT
 26146 FT
 26184 FT
 26222 FT
 26260 FT
 26298 FT
 26336 FT
 26374 FT
 26412 FT
 26450 FT
 26488 FT
 26526 FT
 26564 FT
 26602 FT
 26640 FT
 26678 FT
 26716 FT
 26754 FT
 26792 FT
 26830 FT
 26868 FT
 26906 FT
 26944 FT
 26982 FT
 27020 FT
 27058 FT
 27096 FT
 27134 FT
 27172 FT
 27210 FT
 27248 FT
 27286 FT
 27324 FT
 27362 FT
 27400 FT
 27438 FT
 27476 FT
 27514 FT
 27552 FT
 27590 FT
 27628 FT
 27666 FT
 27704 FT
 27742 FT
 27780 FT
 27818 FT
 27856 FT
 27894 FT
 27932 FT
 27970 FT
 28008 FT
 28046 FT
 28084 FT
 28122 FT
 28160 FT
 28198 FT
 28236 FT
 28274 FT
 28312 FT
 28350 FT
 28388 FT
 28426 FT
 28464 FT
 28502 FT
 28540 FT
 28578 FT
 28616 FT
 28654 FT
 28692 FT
 28730 FT
 28768 FT
 28806 FT
 28844 FT
 28882 FT
 28920 FT
 28958 FT
 28996 FT
 29034 FT
 29072 FT
 29110 FT
 29148 FT
 29186 FT
 29224 FT
 29262 FT
 29300 FT
 29338 FT
 29376 FT
 29414 FT
 29452 FT
 29490 FT
 29528 FT
 29566 FT
 29604 FT
 29642 FT
 29680 FT
 29718 FT
 29756 FT
 29794 FT
 29832 FT
 29870 FT
 29908 FT
 29946 FT
 29984 FT
 30022 FT
 30060 FT
 30098 FT
 30136 FT
 30174 FT
 30212 FT
 30250 FT
 30288 FT
 30326 FT
 30364 FT
 30402 FT
 30440 FT
 30478 FT
 30516 FT
 30554 FT
 30592 FT
 30630 FT
 30668 FT
 30706 FT
 30744 FT
 30782 FT
 30820 FT
 30858 FT
 30896 FT
 30934 FT
 30972 FT
 31010 FT
 31048 FT
 31086 FT
 31124 FT
 31162 FT
 31200 FT
 31238 FT
 31276 FT
 31314 FT
 31352 FT
 31390 FT
 31428 FT
 31466 FT
 31504 FT
 31542 FT
 31580 FT
 31618 FT
 31656 FT
 31694 FT
 31732 FT
 31770 FT
 31808 FT
 31846 FT
 31884 FT
 31922 FT
 31960 FT
 31998 FT
 32036 FT
 32074 FT
 32112 FT
 32150 FT
 32188 FT
 32226 FT
 32264 FT
 32302 FT
 32340 FT
 32378 FT
 32416 FT
 32454 FT
 32492 FT
 32530 FT
 32568 FT
 32606 FT
 32644 FT
 32682 FT
 32720 FT
 32758 FT
 32796 FT
 32834 FT
 32872 FT
 32910 FT
 32948 FT
 32986 FT
 33024 FT
 33062 FT
 33100 FT
 33138 FT
 33176 FT
 33214 FT
 33252 FT
 33290 FT
 33328 FT
 33366 FT
 33404 FT
 33442 FT
 33480 FT
 33518 FT
 33556 FT
 33594 FT
 33632 FT
 33670 FT
 33708 FT
 33746 FT
 33784 FT
 33822 FT
 33860 FT
 33898 FT
 33936 FT
 33974 FT
 34012 FT
 34050 FT
 34088 FT
 34126 FT
 34164 FT
 34202 FT
 34240 FT
 34278 FT
 34316 FT
 34354 FT
 34392 FT
 34430 FT
 34468 FT
 34506 FT
 34544 FT
 34582 FT
 34620 FT
 34658 FT
 34696 FT
 34734 FT
 34772 FT
 34810 FT
 34848 FT
 34886 FT
 34924 FT
 34962 FT
 35000 FT
 35038 FT
 35076 FT
 35114 FT
 35152 FT
 35190 FT
 35228 FT
 35266 FT
 35304 FT
 35342 FT
 35380 FT
 35418 FT
 35456 FT
 35494 FT
 35532 FT
 35570 FT
 35608 FT
 35646 FT
 35684 FT
 35722 FT
 35760 FT
 35798 FT
 35836 FT
 35874 FT
 35912 FT
 35950 FT
 35988 FT
 36026 FT
 36064 FT
 36102 FT
 36140 FT
 36178 FT
 36216 FT
 36254 FT
 36292 FT
 36330 FT
 36368 FT
 36406 FT
 36444 FT
 36482 FT
 36520 FT
 36558 FT
 36596 FT
 36634 FT
 36672 FT
 36710 FT
 36748 FT
 36786 FT
 36824 FT
 36862 FT
 36900 FT
 36938 FT
 36976 FT
 37014 FT
 37052 FT
 37090 FT
 37128 FT
 37166 FT
 37204 FT
 37242 FT
 37280 FT
 37318 FT
 37356 FT
 37394 FT
 37432 FT
 37470 FT
 37508 FT
 37546 FT
 37584 FT
 37622 FT
 37660 FT
 37698 FT
 37736 FT
 37774 FT
 37812 FT
 37850 FT
 37888 FT
 37926 FT
 37964 FT

```
FT FT DISULFID 455 466 BY SIMILARITY.
FT FT DISULFID 460 475 BY SIMILARITY.
FT FT DISULFID 477 486 BY SIMILARITY.
FT FT DISULFID 493 504 BY SIMILARITY.
FT FT DISULFID 498 513 BY SIMILARITY.
FT FT DISULFID 515 524 BY SIMILARITY.
FT FT DISULFID 531 542 BY SIMILARITY.
FT FT DISULFID 536 551 BY SIMILARITY.
FT FT DISULFID 553 562 BY SIMILARITY.
FT FT DISULFID 569 579 BY SIMILARITY.
FT FT DISULFID 574 588 BY SIMILARITY.
FT FT DISULFID 590 599 BY SIMILARITY.
FT FT DISULFID 606 617 BY SIMILARITY.
FT FT DISULFID 611 626 BY SIMILARITY.
FT FT DISULFID 628 637 BY SIMILARITY.
FT FT DISULFID 644 654 BY SIMILARITY.
FT FT DISULFID 649 663 BY SIMILARITY.
FT FT DISULFID 665 674 BY SIMILARITY.
FT FT DISULFID 681 692 BY SIMILARITY.
FT FT DISULFID 686 701 BY SIMILARITY.
FT FT DISULFID 703 712 BY SIMILARITY.
FT FT DISULFID 719 729 BY SIMILARITY.
FT FT DISULFID 724 738 BY SIMILARITY.
FT FT DISULFID 740 749 BY SIMILARITY.
FT FT DISULFID 756 767 BY SIMILARITY.
FT FT DISULFID 761 776 BY SIMILARITY.
FT FT DISULFID 778 787 BY SIMILARITY.
FT FT DISULFID 794 805 BY SIMILARITY.
FT FT DISULFID 799 814 BY SIMILARITY.
FT FT DISULFID 816 825 BY SIMILARITY.
FT FT DISULFID 832 843 BY SIMILARITY.
FT FT DISULFID 837 854 BY SIMILARITY.
FT FT DISULFID 856 865 BY SIMILARITY.
FT FT DISULFID 872 883 BY SIMILARITY.
FT FT DISULFID 877 892 BY SIMILARITY.
FT FT DISULFID 894 903 BY SIMILARITY.
FT FT DISULFID 910 921 BY SIMILARITY.
FT FT DISULFID 915 930 BY SIMILARITY.
FT FT DISULFID 932 941 BY SIMILARITY.
FT FT DISULFID 986 997 BY SIMILARITY.
FT FT DISULFID 991 1006 BY SIMILARITY.
FT FT DISULFID 1008 1017 BY SIMILARITY.
FT FT DISULFID 1024 1035 BY SIMILARITY.
FT FT DISULFID 1029 1044 BY SIMILARITY.
FT FT DISULFID 1046 1055 BY SIMILARITY.
FT FT DISULFID 1062 1073 BY SIMILARITY.
FT FT DISULFID 1067 1082 BY SIMILARITY.
FT FT DISULFID 1084 1093 BY SIMILARITY.
FT FT DISULFID 1100 1121 BY SIMILARITY.
FT FT DISULFID 1115 1130 BY SIMILARITY.
FT FT DISULFID 1132 1141 BY SIMILARITY.
FT FT DISULFID 1148 1159 BY SIMILARITY.
FT FT DISULFID 1153 1168 BY SIMILARITY.

Query Match 2.98; Score 428.5; DB 1; Length 2524;
Best Local Similarity 28.18; Pred. No. 4.7e-15;
Matches 123; Conservative 42; Mismatches 137; Indels 135; Gaps 26;

Qy 532 VQDC-PRNCHGECVSGV-----CHCPFGUGADCAKAAACPVL---CSGNGQYS----- 577
Db 566 INECIPDPCH-YGCKDGIATFTCLCRPGYGRLCDNDINCLSKPCLNGGQCTDRENGY 624
Qy 578 -----KGT-----CQCYSGWKGAECVPMNQCIDP 602
Db 625 ICTCPKGTGVNCEKTIIDCCASNLCDNGKCIDKIDGVECTCEPGYTGKLCININECDSN 684
Qy 603 SCGGHGSC---IDG-NCVCSAGYKGEHC--EEVDCLDPTCSSHGVC---VNG-BCLCSPG 652
Db 685 PCRNGGTCKDQINGFTVCVCPDGYHDHMCLESEVNECNSNPC-IHGACHDGVNGYKCDCEAG 743
Qy 653 WGLNCELARVQC-PDQCSGHGTLYLPTDG--LCSCDPNMWGPDCSVEV--CSVD-CGTHG 706
Db 744 WGSNCNDINNCEBSPCMNGGTCKDMTGAYICTCKAGFGSPNCOTNINECSNPNCLNHG 803
```

```
Qy 707 VCI----GGACRCBEGWTGAACDQVCHP-----RCIEHGTCKDGK-----CECREGNG 752
Db 804 TCIDDVAGYKNCMLPYTGAIC-EAVLAPCAGSPCKNGGRCKESEDFTFSCCEPPGHQ 862
Qy 753 EHTIG-----ROTAGT-----ETDGC-PDLCHNGRGT 780
Db 863 QTCEIDMNECVNRPCRGATCQNTNGSYKCKPKGYTGRNCMDITDDCQPNCHNGSGS 922
Qy 781 LGONSWQCVCTGVRGPGCCNVAMETSCADNKDNEGDLVDCLD----- 823
Db 923 DGINFFCNCBPAGRPGRCEEDI-NECASNPCKNGANTDCVNSYTCYTCQPGFSGIHCS 981
Qy 824 --PDCCLOSAQNSLLC 838
Db 982 NTPD-CTESSCFNGTC 997

RESULT 11
CRB_DROME
ID CRB_DROME STANDARD; PRT; 2139 AA.
AC P10040:
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 01-MAY-2002 (Rel. 41, Last annotation update)
DE Crumbs protein precursor (95F).
GN CRB.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OREGON-R; TISSUE=Embryo;
RA MEDLINE=90263104; PubMed=2344615;
RA Tepass U., Theres C., Knust E.;
RT "Crumbs encodes an EGF-like protein expressed on apical membranes of
RT Drosophila epithelial cells and required for organization of
RT epithelia.";
RL Cell 61:787-799(1990).
RN [2]
RP SEQUENCE OF 1663-1955 FROM N.A.
RC TISSUE=Embryo;
RA MEDLINE=87218537; PubMed=3107986;
RA Knust E., Dietrich U., Tepass U., Bremer K.A., Weigel D.,
RA Vaessin H., Campos-Ortega J.A.;
RT "EGF homologous sequences encoded in the genome of Drosophila
RT melanogaster, and their relation to neurogenic genes.";
RL EMBO J. 6:761-766(1987).
CC -1- FUNCTION: MAY PLAY A ROLE IN THE DEVELOPMENT OF EPITHELIA,
CC POSSIBLY FOR THE ESTABLISHMENT AND/OR MAINTENANCE OF CELL
CC POLARITY. IT MAY ACT AS A SIGNAL.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- PM: PHOSPHORYLATED IN THE CYTOPLASMIC DOMAIN (POTENTIAL).
CC -1- SIMILARITY: CONTAINS 29 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 3 LAMININ G-LIKE DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M33753; AAA28428.1; ALT_SEQ.
DR EMBL; X05144; CAA28793.1; -
DR PIR; B26637; B26637.
DR PIR; A35672; A35672.
DR HSP; P00740; IIXA.
DR FlyBase; FBgn0000368; crb.
DR InterPro; IPR000152; Asx_hydroxyl.
```



```
FT CARBOHYD 550 550 N-LINKED (GLCNAC. . .) (POTENTIAL).
Query Match 2.9%; Score 417; DB 1; Length 2139;
Best Local Similarity 27.1%; Pred. No. 1.5e-14;
Matches 111; Conservative 42; Mismatches 140; Indels 116; Gaps 21;

Qy 529 LDSVQDPRNCHNGECVSGV-----CHCF-PGFIGA-----DCAKAACPVLCSGNGQY 576
Dy 387 VDTDECAQPCQNGSCIDRNGFSCDCSGTGTGAFQCTNVDECDKNPC--LNGGRCLH 444
Qy 577 SKG--TCOCYSRGAECDDPMNOCIDPSCGGHSCIDG-----NCVCSAGYKGEHCEVD 630
Dy 445 TYGWYTCOCLDGMGEICDRPMT-CQTOQCFNGGTCLDKPFGQCLCPPEYTGCLQIAP 503
Qy 631 CLDPTCSSHGVCVNGECLCSFGWGLNCE-----LARVQC-----PDQCS 670
Dy 504 SCAQCPIDSECVGKCKVCKPSSGYNQCTSTGDSALALTPINCNATNGKCLNGGTCS 563
Qy 671 GHGTYLPDTGLCSDDPNMNGPDC-SVEVCS-VDCGTHGVCV----- 709
Dy 564 MNGTH-----CYCAVGYSGDRCEKAENCSPLNCQEPMVCVQNCQCPENKVCNQCATQP 617
Qy 710 --CG-----ACRCEGWTGAC-----DQVCHPCIEHCTCKDG----KCECRE 748
Dy 618 CONGECVDLPNGVECKCTGWTGRTCGNDVDECTLHPKTCGNGICKNEKSGYKCYCTP 677
Qy 749 GWNGEHCTIGRTAGTETDGCPLD-CNGNGRCTLGQNSWCVCQCTGWRGPGCNVAMETSC 807
Dy 678 GFTGVHC-----DSVDDECLSPCLNGATCHKNINAYECVQCPGYEGENCEVDID-EC 729
Qy 808 ADNKNDEGDLVD-----C-LPDCCCLQACQNSLLIC 838
Dy 730 GSNPCNSNGSTCIDRINFTNCIPGMRGRICDIDDDCGVDPCLNGGQC 778

RESULT 12
NOTCH_BRARE
ID NOTCH_BRARE STANDARD; PRT; 2437 AA.
AC P46530;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Neurogenic locus notch homolog protein precursor.
GN NOTCH OR NOTCH1A.
OS Brachydanio rerio (zebrafish) (zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC Cypriniformes; Cyprinidae; Danio.
NCBI_TaxID=7955;
[1]
SEQUENCE FROM N.A.
RN TISSUE=Embryo;
RC MEDLINE=94128602; PubMed=8297791;
RA Blerkamp C., Campos-Ortega J.A.;
RT "A zebrafish homologue of the Drosophila neurogenic gene Notch and
its pattern of transcription during early embryogenesis.";
RL Mech. Dev. 43:87-100(1993).
CC -1- FUNCTION: IMPLICATED IN CELL FATE SPECIFICATIONS DURING
CC EMBRYO DEVELOPMENT. MAY BE INVOLVED IN THE FORMATION OF THE
CC NEURAL PLATE, NOTOCHORD AND BRAIN VESICLES.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED IN ALL CELLS IN PREGASTRULATION
CC STAGES. DURING GASTRULATION IS DIFFERENTIALLY EXPRESSED.
CC ACCUMULATING PREDOMINANTLY IN THE PRECHORDAL MESODERM AND
CC NOTOCHORD. AT THE END OF GASTRULATION, EXPRESSED ALONG THE
CC ANTERIOR-POSTERIOR AXIS INCLUDING THE DEVELOPING NEURAL PLATE
CC AND DIFFERENTIATING MESODERM. ALSO PRESENT IN THE DEVELOPING
CC BRAIN AND HEAD REGIONS.
CC -1- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
CC -1- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC -1- SIMILARITY: CONTAINS 6 ANK REPEATS.
-----
```

```
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
-----
DR EMBL; X69088; CAA48831.1; -.
DR HSP; P00740; 1EDM.
DR ZFIN; ZDB-GENE-990415-173; notch1a.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR001438; EGF_II.
DR InterPro; IPR000800; Notch.
DR Pfam; PF00023; ank; 6.
DR Pfam; PF00008; EGF; 36.
DR Pfam; PF00066; notch; 3.
DR PRINTS; PR00010; EGFBL00D.
DR PRINTS; PR01452; NOTCH.
DR SMART; SM00248; ANK; 5.
DR SMART; SM00179; EGF_CA; 19.
DR SMART; SM00001; EGF_like; 16.
DR SMART; SM00004; NL; 3.
DR PROSITE; PS50088; ANK_REPEAT; 4.
DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
DR PROSITE; PS00010; ASX_HYDROXYL; 23.
DR PROSITE; PS00022; EGF_1; 34.
DR PROSITE; PS01186; EGF_2; 28.
DR PROSITE; PS01187; EGF_CA; 22.
KW Differentiation; Neurogenesis; Repeat; ANK repeat; EGF-like domain;
KW Transmembrane; Signal; Glycoprotein.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 2437 NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN.
FT DOMAIN 21 1724 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1725 1747 POTENTIAL.
FT DOMAIN 1748 2437 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 21 57 EGF-LIKE 1.
FT DOMAIN 58 98 EGF-LIKE 2.
FT DOMAIN 101 138 EGF-LIKE 3.
FT DOMAIN 139 175 EGF-LIKE 4.
FT DOMAIN 177 215 EGF-LIKE 5.
FT DOMAIN 217 254 EGF-LIKE 6.
FT DOMAIN 256 292 EGF-LIKE 7.
FT DOMAIN 294 332 EGF-LIKE 8.
FT DOMAIN 334 370 EGF-LIKE 9.
FT DOMAIN 371 409 EGF-LIKE 10.
FT DOMAIN 411 449 EGF-LIKE 11.
FT DOMAIN 451 487 EGF-LIKE 12.
FT DOMAIN 489 524 EGF-LIKE 13.
FT DOMAIN 526 562 EGF-LIKE 14.
FT DOMAIN 564 599 EGF-LIKE 15.
FT DOMAIN 601 637 EGF-LIKE 16.
FT DOMAIN 639 674 EGF-LIKE 17.
FT DOMAIN 676 712 EGF-LIKE 18.
FT DOMAIN 714 749 EGF-LIKE 19.
FT DOMAIN 751 787 EGF-LIKE 20.
FT DOMAIN 789 825 EGF-LIKE 21.
FT DOMAIN 827 865 EGF-LIKE 22.
FT DOMAIN 867 903 EGF-LIKE 23.
FT DOMAIN 905 941 EGF-LIKE 24.
FT DOMAIN 943 979 EGF-LIKE 25.
FT DOMAIN 981 1017 EGF-LIKE 26.
FT DOMAIN 1019 1055 EGF-LIKE 27.
FT DOMAIN 1057 1093 EGF-LIKE 28.
FT DOMAIN 1095 1141 EGF-LIKE 29.
FT DOMAIN 1143 1179 EGF-LIKE 30.
FT DOMAIN 1181 1217 EGF-LIKE 31.
FT DOMAIN 1219 1263 EGF-LIKE 32.
FT DOMAIN 1265 1303 EGF-LIKE 33.
```

CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).

RA Adachi H., Tsujimoto M., Arai H., Inoue K.;
 RT "Expression cloning of a novel scavenger receptor from human
 RT endothelial cells."
 RL J. Biol. Chem. 272:31217-31220(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Bone marrow;
 RX MEDLINE=96127530; PubMed=8590280;
 RA Nagase T., Seki N., Tanaka A., Ishikawa K.-I., Nomura N.;
 RT "Prediction of the coding sequences of unidentified human genes. IV.
 RT the coding sequences of 40 new genes (K1AA0121-K1AA0160) deduced by
 RT analysis of cDNA clones from human cell line KG-1."
 RL DNA Res. 2:167-174(1995).
 CC -1- FUNCTION: Mediates the binding and degradation of acetylated low
 CC density lipoprotein (Ac-LDL).
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein (Potential).
 CC -1- TISSUE SPECIFICITY: Endothelial cells.
 CC -1- SIMILARITY: CONTAINS 6 EGF-LIKE DOMAINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; D86864; BAA24070.1; -;
 DR EMBL; D63483; BAA09770.1; -;
 DR HSPF; P01180; 2BN2.
 DR InterPro; IPR000561; EGF-like.
 DR SMART; SM00181; EGF; 5.
 DR SMART; SM00001; EGF-like; 2.
 DR PROSITE; PS00022; EGF_1; 6.
 DR PROSITE; PS01186; EGF_2; 6.
 KW Receptor; Glycoprotein; Signal; Transmembrane; EGF-like domain;
 Repeat.
 FT SIGNAL 1 19
 FT CHAIN 20 830
 FT DOMAIN 20 421
 FT TRANSMEM 422 442
 FT DOMAIN 443 830
 FT DOMAIN 53 87
 FT DOMAIN 95 130
 FT DOMAIN 155 191
 FT DOMAIN 215 249
 FT DOMAIN 302 339
 FT DOMAIN 351 382
 FT DOMAIN 476 620
 FT DOMAIN 622 798
 FT DOMAIN 431 438
 FT DISULFID 57 69
 FT DISULFID 63 75
 FT DISULFID 77 86
 FT DISULFID 99 111
 FT DISULFID 105 118
 FT DISULFID 120 129
 FT DISULFID 159 172
 FT DISULFID 165 179
 FT DISULFID 181 190
 FT DISULFID 219 230
 FT DISULFID 225 237
 FT DISULFID 239 248
 FT DISULFID 306 319
 FT DISULFID 313 326
 FT DISULFID 329 338
 FT DISULFID 355 363
 FT DISULFID 358 370
 FT DISULFID 372 381
 FT CARBOHYD 289 289
 FT CARBOHYD 382 382
 FT CARBOHYD 393 393
 FT CONFLICT 662 662
 R -> W (IN REF. 2).

SQ SEQUENCE 830 AA; 87430 MW; F560D9E1AA64D779 CRC64;
 Query Match 2.8%; Score 415.5; DB 1; Length 830;
 Best Local Similarity 25.5%; Pred. No. 4.1e-15;
 Matches 138; Conservative 49; Mismatches 189; Indels 165; Gaps 28;
 QY 533 QDC-----PRNCHNGGECYS-GVCHCFPGFLGADCA-----KAACPVLCSG 572
 DB 50 QECITPICEGPDACOKDEVKVGKGLCRCKPGFFGAHCSRRCPGOYWGPDRCRSCPHPG 109
 QY 573 NGQYSKGTCCQYSGWKGAEDVPMNQCIDPSCGHGSC--IDGNVCVSAGYKGEHVED 630
 DB 110 QCEPATGACQCAQADRWGRCRCEFP---C---ACGPHGRCDPATGVCHCEPGWMSSTCR-P 162
 QY 631 CLDPTCSSHGVCVNGECLCSPGWGLNCELARVQCPCDSCGHGTYL-PDTGLCSGCDPMNM 689
 DB 163 CQNTAAARCEQATGACVCKPQWGRRCSP-----RCNCHGSPCEDSGRCACRPGMW 215
 QY 690 GPDGSVEV-----CSVDC--GTHGVCIGGAC----- 713
 DB 216 GPECQOQCECYGRCSAASGECTCPPGFRGARCELPCPAGSHGVQCAHSCGRCKHNEPCS 275
 QY 714 -----RCREGWTGAACDQ-----RVCHPRC--IEHG-TCK--DGKCE-CREGWNG 752
 DB 276 PDTGSCSECEPGWNGTCCQCPCLPGTFGESECEQCPCHRGHEACRPTDGTGHCQRCDPGWLG 335
 QY 753 EHC-----TIGRQTAGTETDGGPDLCNGNGRCRTLGQNSWQCVCOTGWRGPGCNVAMET 805
 DB 336 PRCEDPCPTGTFEGDCGST-----CPTCVQSGDVTVG-----DCVCSAGYGPSCNACPA 387
 QY 806 SCADNKNE-----GDGLVDCLDPCCLQSAQNSLLRCGRSDPLDIIQOGOT----- 853
 DB 388 GFHGNCSVPCEPEGLCHPVSGCQPCGSRDPTALIVSLVPLLLFLGLACCACCCWA 447
 QY 854 -----DMPAVK-SFYDRIKL-----LAGKDS-----HIIP-----GENPNS 885
 DB 448 PRSOLKDRPADGATVSRMKLVQVWGTLTSLGSTLPCRSLSHKLPWVTVSHHDPEVPFNH 507
 QY 886 SLVSLIRQVVTDTGTPLVGNVSPVKPYG-----YTTITRODGTFDLIANGASIT 938
 DB 508 SFIEPPSAGWATDD-----SFSDDPESGEADEVPAYCVPPQEGMVPVNAQGSSEAS 558
 QY 939 L 939
 DB 559 L 559
 RESULT 14
 NOTC_DROME STANDARD; PRT; 2703 AA.
 AC P07207; P04154; O97458; Q9W4T8;
 DT 01-NOV-1986 (Rel. 03, Created)
 DT 01-MAR-2002 (Rel. 41, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Neurogenic locus Notch protein precursor.
 GN N OR EG:140G11.1 OR EG:163A10.2 OR CG3936.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Oregon-R; TISSUE=Embryo;
 RX MEDLINE=86079539; PubMed=3935325;
 RA Wharton K.A., Johansen K.M., Xu T., Artavanis-Tsakonas S.;
 RT "Nucleotide sequence from the neurogenic locus notch implies a gene
 RT product that shares homology with proteins containing EGF-like
 RT repeats."
 RL Cell 43:567-581(1985).
 RN [2]
 RP SEQUENCE FROM N.A.

RC STRAIN=Canton-S, and Oregon-R; TISSUE=Embryo;
RX MEDLINE=87064624; PubMed=3097517;
RA Kidd S., Kelley M.R., Young M.W.;
RT "Sequence of the notch locus of *Drosophila melanogaster*: relationship
of the encoded protein to mammalian clotting and growth factors.";
RL Mol. Cell. Biol. 5:3094-3108(1986).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN=Berkeley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalili M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kimons I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhou W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of *Drosophila melanogaster*.";
RL Science 287:2185-2195(2000).
RN [4]
RP SEQUENCE FROM N.A.
RX STRAIN=Oregon-R;
RX MEDLINE=20196011; PubMed=10731137;
RA Benos P.V., Gatt M.K., Ashburner M., Murphy L., Harris D.,
RA Barrell B.G., Ferraz C., Vidal S., Brun C., Demailles J., Cadieu E.,
RA Dreano S., Gloux S., Lelaure V., Mottier S., Galibert F., Borkova D.,
RA Minana B., Kafatos F.C., Louis C., Siden-Kimons I., Bolshakov S.,
RA Papogiannakis G., Spanos L., Cox S., Madueno E., de Pablos B.,
RA Modolell J., Peter A., Schoettler P., Werner M., Mourkioti F.,
RA Beinert N., Dowe G., Schaefer U., Jaecckle H., Bucheton A.,
RA Callister D.M., Campbell L.A., Darlamitsou A., Henderson N.S.,
RA McMillan P.J., Salles C., Tait E.A., Valenti P., Saunders R.D.C.,
RA Glover D.M.;
RT "From sequence to chromosome: the tip of the X chromosome of *D. melanogaster*.";
RL Science 287:2220-2222(2000).
RN [5]
RP SEQUENCE OF 2505-2611 FROM N.A.
RX MEDLINE=85099329; PubMed=2981631;
RA Wharton K.A., Redvick B., Finnerty V.G., Artavanis-Tsakonas S.;
RT "opa: a novel family of transcribed repeats shared by the Notch locus
and other developmentally regulated loci in *D. melanogaster*.";
RL Cell 40:55-62(1985).
RN [6]

RP SEQUENCE OF 1-8 FROM N.A.
RX MEDLINE=87257846; PubMed=3037327;
RA Kelley M.R., Kidd S., Berg R.L., Young M.W.;
RT "Restriction of p-element insertions at the Notch locus of *Drosophila melanogaster*.";
RL Mol. Cell. Biol. 7:1545-1548(1987).
RN [7]
RP REVIEW.
RX Harris W.A.;
RT "Many cell types specified by Notch function.";
RL Curr. Biol. 1:120-122(1991).
CC -!- FUNCTION: NOTCH PROTEIN IS ESSENTIAL FOR PROPER DIFFERENTIATION OF
ECTODERM.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- MISCELLANEOUS: SEPARATION OF NEUROBLASTS FROM THE ECTODERM INTO
THE INNER PART OF EMBRYO IS ONE OF THE FIRST STEPS OF CNS
DEVELOPMENT IN INSECTS, THIS PROCESS IS UNDER CONTROL OF THE
NEUROGENIC GENES.
CC -!- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
CC -!- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC -!- SIMILARITY: CONTAINS 6 ANK REPEATS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; M16152; AAB59220.1; -
EMBL; M16153; AAB59220.1; JOINED.
EMBL; M16149; AAB59220.1; JOINED.
EMBL; M16150; AAB59220.1; JOINED.
EMBL; M16151; AAB59220.1; JOINED.
EMBL; K03508; AAB28725.1; -
EMBL; M13889; AAB28725.1; JOINED.
EMBL; K03507; AAB28725.1; JOINED.
EMBL; AE003426; AAF45848.2; -
EMBL; AL035436; CAB37610.1; -
EMBL; AL035395; CAB37610.1; JOINED.
EMBL; M12175; AAA74496.1; -
EMBL; M16025; AAB28726.1; -
PIR; A24420; A24420.
PIR; A24768; A24768.
PIR; A05267; A05267.
HSP; P00740; IIXA.
FlyBase; FBgn0004647; N.
InterPro; IPR002110; ANK.
InterPro; IPR000152; ASX_hydroxyl.
InterPro; IPR000561; EGF-like.
InterPro; IPR000742; EGF 2.
InterPro; IPR001881; EGF_Ca.
InterPro; IPR001438; EGF_II.
InterPro; IPR000800; Notch.
Pfam; PF00023; ank; 6.
Pfam; PF00008; EGF; 36.
Pfam; PF00066; notch; 3.
PRINTS; PR00010; EGFBL00D.
PRINTS; PR01452; NOTCH.
SMART; SM00248; ANK; 4.
SMART; SM00179; EGF_CA; 23.
SMART; SM00001; EGF_like; 13.
SMART; SM00004; NL; 2.
PROSITE; PS00088; ANK_REPEAT; 5.
PROSITE; PS0297; ANK_REPEAT_REGION; 1.
PROSITE; PS00010; ASX_HYDROXYL; 22.
PROSITE; PS00022; EGF_1; 34.
PROSITE; PS01186; EGF_2; 28.
PROSITE; PS01187; EGF_CA; 22.
KW Differentiation; Neurogenesis; Repeat; ANK repeat; EGF-like domain;
Transmembrane; Signal; Glycoprotein.

FT	SIGNAL	1	44	POTENTIAL.
FT	CHAIN	45	1703	NEUROGENIC LOCUS NOTCH PROTEIN.
FT	DOMAIN	45	1745	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	1746	1766	POTENTIAL.
FT	DOMAIN	1767	2703	CYTOPLASMIC (POTENTIAL).
FT	DOMAIN	58	95	EGF-LIKE 1.
FT	DOMAIN	96	136	EGF-LIKE 2.
FT	DOMAIN	139	176	EGF-LIKE 3.
FT	DOMAIN	177	215	EGF-LIKE 4.
FT	DOMAIN	217	253	EGF-LIKE 5.
FT	DOMAIN	255	291	EGF-LIKE 6.
FT	DOMAIN	293	329	EGF-LIKE 7.
FT	DOMAIN	331	370	EGF-LIKE 8.
FT	DOMAIN	372	408	EGF-LIKE 9.
FT	DOMAIN	409	447	EGF-LIKE 10.
FT	DOMAIN	449	486	EGF-LIKE 11.
FT	DOMAIN	488	524	EGF-LIKE 12.
FT	DOMAIN	526	562	EGF-LIKE 13.
FT	DOMAIN	564	600	EGF-LIKE 14.
FT	DOMAIN	602	637	EGF-LIKE 15.
FT	DOMAIN	639	675	EGF-LIKE 16.
FT	DOMAIN	677	713	EGF-LIKE 17.
FT	DOMAIN	715	751	EGF-LIKE 18.
FT	DOMAIN	753	789	EGF-LIKE 19.
FT	DOMAIN	791	827	EGF-LIKE 20.
FT	DOMAIN	829	865	EGF-LIKE 21.
FT	DOMAIN	867	905	EGF-LIKE 22.
FT	DOMAIN	907	944	EGF-LIKE 23.
FT	DOMAIN	946	982	EGF-LIKE 24.
FT	DOMAIN	984	1020	EGF-LIKE 25.
FT	DOMAIN	1022	1058	EGF-LIKE 26.

Query Match		2.8%;	Score 413;	DB 1;	Length 2703;
Best Local Similarity		27.6%;	Pred. No. 3.7e-14;		
Matches 112;		Conservative	45;	Mismatches 143;	Indels 106; Gaps 21;
Qy	531	SVQDCPRN-CHNGECVSGV	-----CHCPGFLGADCAK-----	AACPVLCSGNGQYSGKT	580
Db	602	NINDCDSNPCH-RGKCIDDVANSFRCLCDPGYTGYTCQKQINECESNPQCQFDGHCQDRVGS	660		
Qy	581	--CCYSGMKAEDVPMNQCIDPSCGGHSCIDG----	NCVCSAGYKGEHCEE-VD-CL	632	
Db	661	YYCQCQAGTSKNGEVNVNCHSNPCNNGATCIDGINSYKQCQVPGFTGGQHCERNVDSCI	720		
Qy	633	DPTCSSHGVC---VNGE-----	-----CLCS	650	
Db	721	SSPCANNVCIDQVNGYKCECPRGFYDAHCLSDVDECAASNPCVNEGRCEGDI NEFICHCP	780		
Qy	651	PGWGLNCELARVQCPDQCSHGHTYLPD---	TGLCSDPNMNGPDC--SVEVCSDV-CGT	704	
Db	781	PGYTGKRCLEIDECSSNPCHGGTCTYDKLNAFSCQCMPGYTGQKCEINDDCVTNPCGN	840		
Qy	705	HGVC1---GGACRCEEGWTGAACDORV---	CHPRCIEHGTC-----DGKCECREGWN	751	
Db	841	GGTCLDKVNGYKCVKVPFTGRDESKMDPCASNRCKNEAKCTPSSNFDSCCTCKLGYT	900		
Qy	752	GEHCTIGROTAGTETDGC--	PDLCNGNGRCTLGONSWCVCOTQWRGPGCNVAMETSCAD	809	
Db	901	GRYC-----DEDIDECSLSPCRNGASCLNVPGSYRCCTKGTGEGRDCAINDDCASF	953		
Qy	810	NKDNEG---DGLVD--CL-----	-----DPDCCLQSACONSLLC	838	
Db	954	PCQNGGTCLDGDYSLCVDFGDKHCEETDINECLSQPCQNGATC	999		
RESULT 15					
GLP1_CAEEL		STANDARD;	PRT;	1295	AA.
AC	P13508;				
DT	01-JAN-1990 (Rel. 13, Created)				
DT	01-JAN-1990 (Rel. 13, Last sequence update)				
DT	01-MAR-2002 (Rel. 41, Last annotation update)				
DE	Glpl-1 protein precursor.				

GN	GLP-1 OR EMB-33 OR F02A9.6.
OS	Caenorhabditis elegans.
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC	Rhabditidae; Peloderinae; Caenorhabditis.
OX	NCBI_TaxID=6239;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=BRISTOL N2;
RX	MEDLINE=89336787; PubMed=2758466;
RA	Yochem J., Greenwald I.;
RT	"glp-1 and lin-12, genes implicated in distinct cell-cell
RT	interactions in C. elegans, encode similar transmembrane proteins.";
RL	Cell 58:553-563(1989).
RN	[2]
RP	SEQUENCE FROM N.A.
RC	STRAIN=BRISTOL N2;
RX	MEDLINE=94150718; PubMed=7906398;
RA	Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA	Benfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA	Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
RA	Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA	Johnston L., Jones M., Kershaw J., Kirsten J., Laisster N.,
RA	Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA	Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
RA	Sims M., Smaildon N., Smith A., Smith M., Sonhammer E., Staden R.,
RA	Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA	Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,
RA	Wohlman P.;
RT	"2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT	elegans.";
RL	Nature 368:32-38(1994).
RN	[3]
RP	DELETION OF 1174-1295.
RC	MEDLINE=91351288; PubMed=1881436;
RA	Mango S.E., Maine E.M., Kimble J.;
RT	"Carboxy-terminal truncation activates glp-1 protein to specify
RT	uvulval fates in Caenorhabditis elegans.";
RL	Nature 352:811-815(1991).
RN	[4]
RP	CHARACTERIZATION OF FUNCTION OF THE ANK-REPEATS.
RC	MEDLINE=93354444; PubMed=8350921;
RA	Roehl H., Kimble J.;
RT	"Control of cell fate in C. elegans by a GLP-1 peptide consisting
RT	primarily of ankyrin repeats.";
RL	Nature 364:632-635(1993).
RN	[5]
RP	FUNCTION.
RC	MEDLINE=94208066; PubMed=8156602;
RA	Mello C.C., Draper B.W., Priess J.R.;
RT	"The maternal genes apx-1 and glp-1 and establishment of
RT	dorsal-ventral polarity in the early C. elegans embryo.";
RL	Cell 77:95-106(1994).
CC	-1- FUNCTION: INVOLVED IN THE SPECIFICATION OF THE CELL FATES OF THE
CC	BLASTOMERES ABA AND APA. PROPER SIGNALING BY GLP-1 INDUCES ABA
CC	DESCENDANTS TO PRODUCE ANTERIOR PHARYNGEAL CELLS, AND APA
CC	DESCENDANTS TO ADOPT A DIFFERENT FATE. CONTRIBUTES TO THE
CC	ESTABLISHMENT THE DORSAL-VENTRAL AXIS IN EARLY EMBRYOS.
CC	-1- SUBCELLULAR LOCATION: Type I membrane protein.
CC	-1- DEVELOPMENTAL STAGE: ACTS ON ABP DEVELOPMENT DURING 4-CELL AND
CC	12-CELL STAGES, AND ON ABA DEVELOPMENT DURING 12-CELL AND 28-CELL
CC	STAGES.
CC	-1- SIMILARITY: HIGH, TO C.ELEGANS LIN-12.
CC	-1- SIMILARITY: CONTAINS 10 EGF-LIKE DOMAINS.
CC	-1- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC	-1- SIMILARITY: CONTAINS 5 ANK REPEATS.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC	or send an email to license@isb-sib.ch).

CC EMBL; M25580; AAA28058.1; -
DR EMBL; Z19555; CAA79620.1; -
DR EMBL; Z29116; CAA79620.1; JOINED.
DR EMBL; Z29116; CAA82373.1; -
DR EMBL; Z19555; CAA82373.1; JOINED.
DR PIR; A32901; A32901.
DR HSP; P00740; 1EDM.
DR WormPep; F02A9.6; CE00237.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR000800; Notch.
DR Pfam; PF00023; ank; 6.
DR Pfam; PF00008; EGF; 10.
DR Pfam; PF00066; notch; 3.
DR PRINTS; PR01452; NOTCH.
DR SMART; SM00248; ANK; 4.
DR SMART; SM00179; EGF_Ca; 1.
DR SMART; SM00001; EGF_Like; 7.
DR SMART; SM00004; NL; 3.
DR PROSITE; PS50088; ANK_REPEAT; 3.
DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
DR PROSITE; PS00010; ASX_HYDROXYL; 2.
DR PROSITE; PS00022; EGF_1; 10.
DR PROSITE; PS01186; EGF_2; 8.
DR PROSITE; PS01187; EGF_Ca; 1.
KW Differentiation; Repeat; ANK repeat; EGF-like domain; Transmembrane;
KW Glycoprotein; Signal.
FT SIGNAL 1 15
FT CHAIN 16 1295
FT DOMAIN 16 764
FT TRANSMEM 765 786
FT DOMAIN 787 1295
FT DOMAIN 19 58
FT DOMAIN 117 152
FT DOMAIN 154 190
FT DOMAIN 190 230
FT DOMAIN 232 269
FT DOMAIN 271 308
FT DOMAIN 316 359
FT DOMAIN 369 406
FT DOMAIN 407 443
FT DOMAIN 446 479
FT REPEAT 493 527
FT REPEAT 528 568
FT REPEAT 569 608
FT REPEAT 961 990
FT REPEAT 994 1023
FT REPEAT 1030 1062
FT REPEAT 1074 1103
FT REPEAT 1107 1136
FT DISULFID 23 35
FT DISULFID 29 46
FT DISULFID 48 57
FT DISULFID 121 131
FT DISULFID 126 140
FT DISULFID 142 151
FT DISULFID 158 169
FT DISULFID 163 178
FT DISULFID 180 189
FT DISULFID 194 206
FT DISULFID 201 218
FT DISULFID 220 229
FT DISULFID 236 248
FT DISULFID 242 257
FT DISULFID 259 268
FT DISULFID 275 286
FT DISULFID 280 296
FT DISULFID 298 307
FT DISULFID 373 384

FT DISULFID 378 394 BY SIMILARITY.
FT DISULFID 396 405 BY SIMILARITY.
FT DISULFID 411 422 BY SIMILARITY.
FT DISULFID 416 431 BY SIMILARITY.
FT DISULFID 433 442 BY SIMILARITY.
FT DISULFID 450 461 BY SIMILARITY.
FT DISULFID 455 467 BY SIMILARITY.
FT DISULFID 469 478 BY SIMILARITY.
FT CARBOHYD 244 244 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 245 245 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 333 333 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 381 381 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 609 609 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 675 675 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 1295 AA; 144078 MW; 422AADOA2DEEF3B4 CRC64;

Query Match 2.7%; Score 399; DB 1; Length 1295;
Best Local Similarity 20.3%; Pred. No. 6.7e-14;
Matches 233; Conservative 139; Mismatches 370; Indels 408; Gaps 57;

QY 530 DSVDQCPRN-CHGNBCEVSGV-----CHCFPGFLGADCAKAACPVL-----CSNGOY 576
DB 153 EGDHCAQNECAEGSTCVSNVYVYCDPIGKSGRYCERTECALMGNICNHGRCIPNDE 212
QY 577 SKG-TQCYSYGMKGAEDVPMNQC-IDPSGGHSGCI-----DGNVCYSAGYKGEHCEE-V 629
DB 213 DKNFRVCDSGYEGEFCNKNKNECLTEETCVNNTSFCNLHGDFTCTCKPGYAGKYCEEAI 272
QY 630 D-CLDPTCSSHGVNCE-----CLCSPWGLNCELARVQCPD-----QC 669
DB 273 DMCKDYVCNDGYCAHDSNQMPICYCEQFTGQRC-----IECPGFGGTHCDLPLQRPHC 329
QY 670 S-GHGTPLPD-----TGLCSDPNMMPDC-----SVEVCSVD-CGTHGVCI----- 709
DB 330 SRNGTCYNDGRINGFCVCEPDYIGDCEINRKPDKFPDIOSCKYPCVNNATCIDLN 389
QY 710 -GGACRCBEGWTGAACDQR-VCHP-RCIEHGTC--DG--KCECRGWNGEHCCTI-GROT 761
DB 390 SGYSCHCPLGFYGLNCEQHLCTPTTCANGTCEGVNGVIRNCNPGFSGDYCEIKDROL 449
QY 762 AGTETDGCPLDNGNGRCITLGNOSWCVCOTGWRGPGCN--VAMETS-----CADNK 811
DB 450 CSRHP-----CKNGVC---ANTGYCEQYGTGTCBEVLVIEKSKETVINDLCEQRK 500
QY 812 --DNEGDLVDCIDPDCCIQSACQNSLLRCGRSDPLDIIQOQTDMPA-----857
DB 501 CMDLASNGI--CNPECNLEECNFDGDCSGGQRPFSKCQ-----YPARCAQFANGVCN 552
QY 858 -----VKSFYDRIKLAGKDSHIIIPGENPFSNLSV---SLIRGVVTTDGTPLVGVNVS 909
DB 553 QECNNEECLYD-----GLDC-----QSELCFPAHIRKHCIERRGDGVNCLCS 596
QY 910 FVKYPKYG-----YTITRODGTPLIANGGASL-----TLHEFR---943
DB 597 FICGDFDGGDCNNGTEAIIISDIRIKVQIDPIEFQATGGETLMQISANLRATVRIORDEL 656
QY 944 APFMSQERTVWLPWNSFYAMDITLVKMTENSIPLSCDSGFVRPDPDIIISPLSTFFSAAP 1003
DB 657 GPLVFR-----WDGEHEMERVEMNSSK-----LEDQFVLSHHVRRYRQA--695
QY 1004 GONPIVPTQVLHEEIELPGSNVKLYLSRRTAGYKSLKTKMTQSTVPLNLRVILMVA 1063
DB 696 -----VVTGLVLYLEVE---EICKPEFCRFSTA-----QSVVDL-----726
QY 1064 VEGHLFQKSFQASPNLASTFINDKTDAYGORVYGLSDAVSVGVFEYETCPSLILWEKRTA 1123
DB 727 -----IAAGLV-----KSDGRMSLG-----LPITAMVA 750
QY 1124 LLQGFELPSNLGGWSLKHHL-----NVKSGILHKTGE-----NOFLTQO 1166
DB 751 VPKRNEIDE-----GWSRSQVILLFACIAFLAFGVVAGVIKNGPERSRKRKMVNATVWMP 806

QY 1167 PAITSIMGRRRRSISPCSCNGL-----AEGNKLLAPVALAV 1204
Db 807 PHESTNEKGR-RNQSNHSSQCSSLDNSAYHPNTRKHCSDYSTGYNGEQYSQIYPQTLAN 865
QY 1205 GIDGSLYVGDPNYIRRIIPSRNVTISILELRNKEFKHSNPAHKYYLAVDPVSGSLYVSDT 1264
Db 866 G-----YPGDYNELN-----FDFQSETFAPADLPADLPLHVQAAGPDAITAPI 909
QY 1265 NSRRIYRVKSLSGTKDLAGNSEWAGTGEQCLPFDEARCGDGGKAIDATLMSPRGIAVDK 1324
Db 910 TNESVNQVDSKYRRRVLHWAANVRGKPEDVITTEAIRCLKAGADYNA----- 957
QY 1325 NGLMYFVDATMIRKVDQNGIISTLGSNDLTAVRPLSCDSSMDVAQVRELPDLDVAVNPM 1384
Db 958 -----RDCDENTAL-----MLAVR-----AHRVRL----- 977
QY 1385 DNSLYVLENNVILRITENHOVSIIAGRPMHCQVPGIDYSLSKLAHSALESASATAISHT 1444
978 -----SVLLREGAN-----PTIFNNSERSALHEAVVVKDLRLRH- 1013
QY 1445 GVLYITETDEKKINRLQVTTNG-----EICLLAGAAASDCDCKNDV 1485
Db 1014 -----LLTDKRLKKEIDELDRNGMTALMLVARELKGHVEMAEALLSKGAKLDYDGAARK 1068
QY 1486 NCNCYSGDDA 1495
Db 1069 DSNKYKGRTA 1078

Search completed: September 18, 2002, 11:17:02
Job time: 517 sec

